ABSTRACT

Title of Thesis: ESTIMATING THE VALUE OF PERMANENTLY PRESERVED AND DEVELOPABLE OPEN SPACE IN AN URBAN FRINGE COUNTY

Degree candidate: Shawn Joseph Bucholtz

Degree and year: Master of Science, 2003

Thesis directed by: Professor Loretta M. Lynch
Department of Agricultural and Resource Economics

This thesis investigates how adjacent permanently preserved open space and developable open space affects housing values in an urban fringe county using a hedonic model of housing prices in Montgomery County, Maryland. The value of open space is measured in two areas around a residential parcel: the viewshed and the neighborhood. This thesis investigates three important methodological issues when estimating the value of open space. First, the proposed hedonic model incorporates spatial dependence in the error tem, using a generalized-banded approach to building a more flexible spatial weight matrix, based upon boundary relationships at the city-block and sub-division level. This thesis finds that parameter estimates are robust to changes in the specification of the weight matrix. Second, this thesis tests for sensitivity of the land use parameter estimates to changes in the size of the viewshed and neighborhood. It is found that the land use parameter estimates are robust to changes in the size of the viewshed or neighborhood.

Third, this thesis tests for endogeneity of the land use measures and finds that endogeneity varies depending on the size of the viewshed. An Instrumental Variables approach to estimating the model is used to address the endogeneity issues.

In the final part of this thesis, the land use parameter estimates are used to determine the total change in residential property values from two local conversions of developable open space to permanently preserved open space. It is found that these local conversions increase nearby property values in the viewshed but decrease property values in the neighborhood. In almost all cases, the net change in value was negative.

ESTIMATING THE VALUE OF PERMANETLY PRESERVED AND DEVELOPABLE OPENSPACE IN AN URBAN FRINGE COUNTY

By

Shawn Joseph Bucholtz

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2003

Advisory Committee:

Professor Loretta M. Lynch, Chair
Professor Nancy Bockstael
Dr. Michael Roberts

ACKNOWLEDGEMENTS

I would like to thank my committee members for their hard work in guiding my progress through this thesis. Specifically, I would like to thank Dr. Lori Lynch for her guidance over the past three years and willingness to work with me on several projects. She has shown tremendous patience with me on our joint projects as I attempted to learn a bit about spatial econometrics. My graduate education was a positive experience because of her presence.

I would like to thank Dr. Michael Roberts for his guidance on all things econometrics. I appreciate his ability to present econometrics in an intuitive way, which seems to be the best way for me to figure it out.

I would like to thank Dr. Nancy Bockstael for her wisdom and great ideas for my thesis. I also appreciate her willingness to let me use various data sets.

I would like to give many thanks to my coworkers at the Economic Research Service for their suggestions on this topic and willingness to provide constructive comments. Most of all, I would like to thank my boss, Vince Breneman, for his patience as I completed the work, and for hanging a big raise on the completion of this thesis.

I would like to thank my graduate school friends, especially Glen, Valerie, Mike, Shannon, and Jeremy for their constant encouragement and willingness to listen to me complain about graduate school.

Most importantly, I would like to thank my family for their support over the past three years. I would especially like to thank my parents, Carol, and Herb and Delores, for the encouragement.

TABLE OF CONTENTS

List of Tables vi

List of Figures viii

Chapter I: Introduction 1
The Benefits of Open Space 1
Methodological Issues in Hedonic Models of Residential Parcel Sales 4

Chapter II: The Hedonic Model 7
Overview 7
Assumptions of the Hedonic Model 8

Chapter III: Spatial Effects in Hedonic Models 10
A Non-Spatial, Linear Hedonic Model. 10
Spatial Effects: Spatial Heterogeneity and its Effect on the Hedonic Model 11
Spatial Effects: Spatial Dependence and its Effect on the Hedonic Model 12
The General Spatial Model 13

Chapter IV: The Spatial Weight Matrix 15
Overview 15
Contiguity-Based Spatial Weight Matrices 17
Distance-Based Spatial Weight Matrices 18
Cliff and Ord's Spatial Weight Matrix 20
Spatial Weight Matrices with Estimated Parameters 20
Row Standardization of the Spatial Weight Matrix 21
Bell and Bockstael's Generalized-Banded Spatial Weight Matrix 22

Chapter V: Spatial Dependence and Geographic Boundaries 24
Overview 24
Unobservable Quality of Access to Local Parks 25
Unobservable Quality of Access to Parking 26

Chapter VI: Review of Literature on Hedonic Models 28
Overview 28
Hedonic Models of Locally Undesirable Land Uses and Disamenities 28
Hedonic Models of Locally Desirable Land Uses 30
Hedonic Models of the Value of Open Space in Maryland 30

Chapter VII: The Proposed Hedonic Model 33
Overview 33
The Hedonic Model 34
Functional Form 37
The Weight Matrix for the Error Term 38
Alternative Weight Matrices 39
Potential Endogeneity of Land Uses Indices 41

Chapter VIII: Method of Moments Estimation of Spatial Models 43
Overview 43
Generalized Moments Approach to the Spatial Error Model with One Spatial Weight Matrix 44
Generalized Moments Approach to the Spatial Error Model with Two Spatial Weight Matrices 46

Chapter IX: The Data Sets 49
Overview 49
Housing Sales Data 49
Parcel Boundaries Data Set and Buffers 50
Land Use Data Sets 52
School District Boundaries 54
Soils Data Set 56
Sewer Data 56
Constructing the Spatial Weight Matrices 56
Other Data Sources 57

Chapter X: Exploratory Data Analysis 58
Overview 58
Housing Structure and Neighborhood Variables 66
Land Use Indices 63
The Spatial Weight Matrices 70

Chapter XI: Endogeneity and Spatial Dependence Test Results 72
Tests for Endogeneity 72
Tests for Spatial Dependence 74
Mapping of Errors 76

Chapter XII: Robustness of Results to Changes in Viewshed and Neighborhood Sizes and Spatial Weight Matrices 79
Overview 79
Summarizing the Results of the Structure, Neighborhood, and Sales Period Variables 82
The Robustness of the Land Use Index Parameter Estimates to Changes in the Spatial Weight Matrix Specification 85
The Robustness of the Land Use Index Parameter Estimates to Changes in the Viewshed-Neighborhood Combinations 87

Comparing the Results of the Spatial Dependence Coefficients 89
Conclusions 92

Chapter XIII: Interpreting the Land Use Parameter Estimates and Simulating Localized Increases in Permanently Preserved Open Space 93
Overview 93
Comparing the Land Use Index Parameter Estimates 94
Simulation Process 96
Results of Simulations 99

Chapter XIV: Conclusion 104

References 106

13.3 Net Value of Conversion of 11 Acre Parcel to Permanently Preserved Open Space in Poolesville, Maryland 101

LIST OF FIGURES

4.1	A Typical Lattice Structure	179
4.2	Comparing Top Two Rows of Actual, Inverse, and Row-Standardized Inverse Distance Weights	21
4.3	Bell and Bockstael's Bands	22
5.1	Quality of Access to Local Parks	26
7.1	Land-use Buffers Sizes: Map of Two Viewsheds and Two Neighborhoods	36
7.2	Local Map of Block and Sub-division Neighbors for a Residential Parcel	39
7.3	Local Map of 300 and 300-600 Neighbors for a Residential Parcel	40
9.1	Buffering a Representative Parcel	51
9.2	Six Aggregated Categories of Land Use, Montgomery County	54
9.3	Montgomery County High School Districts, 1998-1999	55
10.1	Histogram of Housing Price with Density Curves	61
10.2	Histogram of Age with Density Curves	62
10.3	Histogram of Square Foot of Structure with Density Curves	62
10.4	Histogram of Lot Size (Acres) with Density Curves	63
10.5	Histogram of Median Household Income with Density Curves	63
10.6	Histogram of Population Density with Density Curves	64
10.7	Histogram of Distance to Washington, DC with Density Curves	64
10.8	Histogram of Distance to Train Stations with Density Curves	65
13.3	Map of OLS Residuals--Germantown Region of Montgomery County	77
13.4	Map of OLS Residuals--Silver Spring Region of Montgomery County	78

Chapter 1

INRODUCTION

1.1 The Benefits of Open Space

The preservation of open space is an important policy issue in the State of Maryland. Open space, such as forest and farmlands, provides recreation opportunities, habitat, aesthetic amenities, and ecosystem services. In most cases, forest and farm land are part of the private land market. When the private land market fails to account for all the externalities associated with forest and farm land, economic theory suggests that market failure has occurred and that the optimal configuration of open space parcels, given public preferences, will not be preserved. The optimal configuration of open space parcels is not one dimensional, nor is it static. As the landscape changes, i.e. the spatial extent, spatial arrangement, and size distribution of parcels of open space, the optimal number and the location of preserved open space parcels will change.

To overcome these economic inefficiencies, private and governmental entities have established several policy tools to preserve open space. For example, in 1977, the State of Maryland established the Maryland Agricultural Land Preservation Foundation (MALPF). This type of program is commonly referred to as a "Purchase of Development Rights" (PDR) or in some circles a "Purchase of Agricultural Conservation Easement" Program (PACE). The funding for MALPF comes from a real estate transfer tax and an agricultural transfer tax. In fiscal year 1999, MALPF spent $23.1 million and counties

increase this total by contributing $6.7 million additional dollars (Chesapeake Bay Commission, 2001). In addition, Montgomery County has introduced a "Transfer of Development Right" program (TDR) to preserve land in its 90,000 Agricultural Reserve. In 1990, it introduced a county-level PDR program to supplement the state program. Together the state and county PDR programs have expended $28 million on easements since 1980. Montgomery County has preserved 51,132 acres of privately owned farm and forest land. Preserved acres were 61.56% of 1997 farmland in the county or 15.98% of the county's land (United States Department of Agriculture, 1997, Bowers 2000). Montgomery has preserved 2,122 acres under Maryland Environmental Trust (MET). In the MET program, a landowner donates the conservation easement he or she has negotiated with MET and receives a tax deduction on his or her federal and state income taxes. Owners typically benefit from estate tax savings as well.

In addition to preserved farmland, Maryland owns and operates many small parks and recreational area and two large parks, each of approximately 6,100 acres, in Montgomery County. In addition, Montgomery County has purchased 30,000 acres which are used as public parks. All told this is 13.19% of Montgomery's 320,000 acres of land. Beyond parks and recreational areas, the Revenue Authority, an instrumentality of the County and a public corporation, owns and operates five 18-hole public golf courses. The county parks and planning department manages five additional golf courses. There are also numerous private golf courses in the county. While golf courses are not a large number of acres, for example the largest one in Poolesville is only 262 acres; they do provide open space in close proximity to urban and suburban areas.

While these acres and expenditures are significant, they may still be insufficient

to preserve the optimal configuration of open space parcels. To achieve the optimal configuration of open space parcels, policy makers need information on the marginal benefits and costs of preserving open space. Because open space is a public good, benefits are realized by the public. Benefits such as aesthetic value may be realized by persons living nearby, while other benefits such as water quality improvement may be accrued by the entire county.

As a beginning to estimating the benefits of open space, this thesis focuses on those benefits of open space accruing to nearby property owners. These benefits may be capitalized into the prices of improved residential parcels (parcels with a housing structure), hereafter referred to as residential parcels as new buyers will pay more for the parcel. If this increase in residential parcel value exists, then it may be recaptured through property taxes. For example, if a residential parcel near a parcel of open space is worth more than a similar residential parcel not in the vicinity of open space, then that open space may be the cause of the increased value. That increased value of the residential parcel typically means a higher taxable value. These additional tax dollars may be source of financing for further open space preservation programs (Geoghegan, Lynch, and Bucholtz, 2003).

Previous research has found mixed results in assessing if open space preservation programs are capitalized into the price of residential parcels. Several studies have found that preserved open space can increase the value of residential parcels (Geoghegan, 2002; Irwin and Bockstael, 2001; Irwin, 2002). Another study found that the value varied from county to county and the size of the radius used to measure the surrounding land use can affect the value of the open space (Geoghegan, Lynch, and Bucholtz, 2003).

This thesis estimates the benefit to residential parcel owners from different types of nearby open space. A hedonic model of residential parcel prices is used to compare the value of open space that will remain open space (preserved open space) to open space that can be converted to residential or commercial uses at some point in the future (developable open space) to nearby owners. This thesis examines if adjacent landowners received benefits from and were willing to pay more because of the adjacent open space and if so, how much more. In addition, the thesis addresses several methodology issues.

1.2 Methodological Issues in Hedonic Models of Residential Parcel Sales

Hedonic analysis, particularly in the housing market, is a tool that is used to conduct cost-benefit analysis on policy decisions. Many of the earlier hedonic models of housing markets ignored issues that may arise from the spatial nature of the data sets which may have resulted in flawed estimates. Addressing the spatial nature of data sets, particularly in the housing sector, is important to ensure correct estimates and sound policy decisions.

Spatial dependence is the issue most frequently discussed in the hedonic literature. Spatial dependence can occur when the price of one residential parcel affects the price of another, in a structural way, or when spatially correlated variables are unobservable to the researcher and thus omitted from the model. To correct for potential spatial dependence, a spatially-lagged dependent variable is included in the model and/or spatial dependence is modeled in the error term. Both of these methods require specifying and using a spatial weight matrix.

The structure or form of the spatial dependence is contained in the spatial weight matrix. Many researchers assumed that it is more consistent with economic behavior that neighbors have differential impacts on each other and that those impacts decline as distance between neighbors increase. Thus, they use distance-based spatial weight matrices. These matrices permit a stronger relationship among neighbors close together than among those neighbors at a greater distance. However, these distance-based spatial weight matrices have some limitations. For example, their structure generates a spatial dependence that varies smoothly across space. Yet, economic theory does not suggest or require that spatial dependence should vary smoothly across space and in fact spatial dependence may not be smooth. It is also a standard practice to establish a cut-off distance, which suggests that relationships exist only up to a certain spatial extent. In addition, many researchers row-standardize the spatial weight matrix, which changes the underlying relationships between neighbors.

Noting these limitations, more flexible forms of distance-based spatial weight matrices have been used to capture spatial dependence in the error terms of hedonic models of residential parcel prices. Motivated mainly by issues regarding row-standardization, Bell and Bockstael (2000) built a generalized weight matrix using three bands (200, 400, and 600 meters), each with their own spatial dependence parameter. Within each band, the spatial dependence was assumed to be equal among all neighbors. Their approach is similar to constructing a contiguity-based spatial weight matrix in which if one is a neighbor the off-diagonal term is one, and if one is not, zero.

This thesis proposes an alternative specification of the form of spatial dependence using a generalized-banded approach to building the spatial weight matrix, motivated by

the research of Bell and Bockstael (2000). However, this thesis deviates from Bell and Bockstael's work by hypothesizing that bands built on contiguity-based relationships within blocks and sub-divisions more accurately reflects the true spatial dependence among residential parcel sales, specifically in the error term of a hedonic model.

In addition to addressing the spatial issues in hedonic models, two other methodological issues are addressed in this thesis. The first issue is endogeneity of the land use measures. The second methodological issue involves the measurement of land use indices. Past research has calculated land use indices based on the amount of various types of land use surrounding a parcel. However, there is no consensus on how large the surrounding area should be. This thesis investigates how different sizes for the surrounding area may yield different parameter estimates for land use indices, and more specifically, permanently preserved open space indices.

The primary objective of this thesis is to estimate the values of different types of open space in the residential parcel market in Montgomery County correctly. The secondary objectives are to determine how alternative specification of the form of spatial dependence, the presence of endogeneity, and different measurement schemes for the land use indices change the parameter estimates for types of open space.

Chapter 2

THE HEDONIC MODEL

2.1 Overview

Hedonic analysis is often used to measure the implicit price of attributes of a particular good. Hedonic analysis posits that the price of a good is a function of its characteristics. It has often been used in the residential parcel market, beginning with Ridker and Henning's (1976) investigation of the relationship between air quality and property values.

The Hedonic Price Function (HPF) is a market clearing function produced by the interaction of bid functions of households and offer functions of suppliers (Rosen, 1974; Can, 1992). The HPF formalizes a relationship between the price of a good and vector of explanatory variables. A typical function of residential parcel values is

$$P = f(S, L, N) \qquad (2.1)$$

where P is a vector of housing prices, S is a matrix of structural characteristics, L is a matrix of surrounding land use characteristics, and N is a matrix of neighborhood characteristics. All else being equal, the partial derivative of the above hedonic function with respect to any attribute is its implicit marginal price (Rosen, 1974). In the linear model, the implicit price of the housing attribute is revealed in the regression coefficient.

The primary advantage of using hedonic models is that they are a market-based, revealed preference approach to the valuation of characteristics of goods. In terms of the residential parcel market, HPF requires information on the price and attributes of the parcel and its improvements, as well as the specification of a functional form.

2.2 Assumptions of the Hedonic Model

The application of the hedonic price model to the housing market rests on a few key assumptions. First, the residential parcel market is assumed to operate under perfectly competitive conditions where there are numerous buyers and sellers. Under this assumption, no individual buyer or seller is able to significantly affect the price of properties through purchases or sales. This assumption may be violated if a small number of sellers account for the majority of the market, such as a few developers who are building many of the new homes in a particular area. However, most hedonic models are estimated on all sales and most sellers will be residential parcel owners. Therefore, this assumption should rarely be violated.

Second, buyers and sellers are assumed to have perfect information. This assumption impacts the buyers in the residential market in two ways. First, the buyer needs to have complete information concerning all possible opportunities to purchase. Most of this information is easily available through media such as newspapers and real estate listings. Buyers are also impacted by the amount and kind of information they acquire on individual units for sale. While most of the relevant information is available from the same sources that list the properties, other pieces of information may not be so

readily available. Many single family homebuyers thus will hire an inspector to investigate a property before they purchase it. The buyer then relies on the inspector's expertise to fill in the information gap between what can be observed with minimal knowledge and perfect information. In addition, regulations have been adopted to improve information homebuyers receive about the property they intend to purchase. A good example is the mandatory reporting of lead paint and radon.

From the prospective of sellers, perfect information would allow them to maximize their profit. Sellers typically hire realtors with extensive information about the market to fill in the information gap between what the owners think the price should be and the true market price.

Third, the HPF model assumes the market is in equilibrium. This assumption holds if the price vector adjusted instantaneously to changes in the supply or demand at any point in time. This assumption is less likely to hold in the short-run because of imperfections in the real world property market. In the long-run, market equilibrium is likely.

In addition, to these basic assumptions about HPF, in estimating regression models, researchers make assumptions. The next chapter discusses some of the impacts of ignoring spatial effects when estimating a HPF.

Chapter 3

SPATIAL EFFECTS IN HEDONIC MODELS

3.1 A Non-Spatial, Linear Hedonic Model

Chapter 2 presented a typical function of residential parcel values where price was a function of structural, land use, and neighborhood characteristics. The linear function can be formalized into the model

$$P = \alpha + S\beta + L\eta + N\varsigma + \epsilon \tag{3.1}$$

And

$$\epsilon \sim N(0,\sigma^2 I),$$

where P is an n x 1 vector of housing prices, S, L, and N are n x k matrices of independent variables, and α, β, η, and ς are vectors of parameters associated with the independent variables. The disturbance ϵ is taken to be normally distributed with constant variance and zero covariance among the error terms.

The model presented in 3.1 does not account for potential spatial effects. There is no spatially-lagged dependent variable and parameters are not allowed to vary across space. In addition, the error terms are identically and independently distributed, not allowing for heteroscedasticity or non-zero covariance.

3.2 Spatial Effects: Spatial Heterogeneity and its Effect on the Hedonic Model

Spatial heterogeneity is a violation of a standard assumption under normality of the error of a linear regression model – identically distributed error terms. It is commonly associated with a misspecification resulting from omitted variables, incorrect functional specification, or measurement errors, and is often treated as a nuisance.

Failing to account for spatial heterogeneity may result in heteroskedasticity. Heteroskedasticity occurs when the variance of the error terms is not constant. When the errors are heteroskedastic, the OLS estimator remains unbiased, but is no longer efficient. More importantly, typical hypothesis testing procedures are no longer appropriate because standard errors are biased (Greene, 2000, pg. 502).

One way spatial heterogeneity may arise is from *neighborhood effects*. Can (1992, pg. 456) defines neighborhood effects as "...a set of characteristics pertaining to the socioeconomic and physical make-up of the neighborhood, accessibility to urban amenities, and the level of public services." The term "neighborhood" is itself not easy to delineate, either conceptually or geographically. It may range from an urban block to a suburban sub-division to a rural town.

In the context of a hedonic model, neighborhood effects may result in the sub-markets within the assumed uniform residential parcel market and may require separate hedonic models for each neighborhood. However, neighborhood effects can be modeled in one hedonic by allowing the neighborhoods to enter as direct determinant (neighborhood dummies) or by allowing some exogenous variables to vary across (interaction of neighborhood dummies and exogenous variables).

3.3 Spatial Effects: Spatial Dependence and its Effect on the Hedonic Model

Like spatial heterogeneity, spatial dependence is also a violation of a standard assumption under normality of the errors of a linear regression model. However, spatial dependence violates the assumption of independently distributed error terms, that is, non-zero covariance between pairs of errors.

Spatial dependence is typically the result of the omissions of two types of variables. The first omission is of a spatially lagged dependent variable. Given that a common practice in real estate is to determine the price of a house based on comparables in the "neighborhood", spatially lagged dependent variables may be important explanatory variables. These spatial lag structures need to incorporate the time of sale. For example, a residential parcel should only include residential parcels that were sold at a prior date as explanatory variables.

The second omission is that of unobserved variables that are spatially correlated. This results in non-zero covariance terms in the variance-covariance matrix of the error term. For example, the levels of air pollution will vary depending on factors such as wind and proximity to pollution sources. The level of air pollution at any given point may be correlated with pollution at other points. The correlation should decrease (increase) between two points as the distance from each other increases (decreases).

The exclusion of a spatially lagged dependent variable amounts to under-specification, which biases the parameter estimates. Anselin (1988, pg. 58) concludes that the inclusion of spatially lagged dependent variables leads to an OLS estimator that

will be biased as well as inconsistent, irrespective of the properties of the error term. Therefore, OLS can not be used if a spatially lagged dependent variable is included.

Failing to model spatial dependence in the error term will lead to parameter estimates that are unbiased, but inefficient, due to the non-diagonal structure of the error covariance matrix. In addition, no consistent estimates for the autoregressive parameter can be obtained from OLS estimation (Anselin, 1988, pg. 59).

3.4 The General Spatial Model

The general spatial model in this section presents a formalization of the three spatial effects discussed in the previous section. The model developed below follows directly from Anselin (1988, pg. 34):

$$\begin{aligned} Y &= \rho WY + X\beta + \epsilon \\ \epsilon &= (I - \lambda W)^{-1}\mu \\ \mu &\sim N(0,\Omega) \\ \Omega_{ii} &= h_i(Z\alpha), \end{aligned} \tag{3.2}$$

where Y is a n x 1 vector of dependent variables (in this case it would be price (P)), ρ is the coefficient of the spatially lagged dependent variable, β is a k x 1 vector of parameters associated with exogenous variables, X is an n x k matrix of exogenous variables (S,L, N from equation [3.1]), ϵ is an n x 1 vector of disturbances, λ is the coefficient of the spatial autoregressive structure for the disturbance ϵ, and μ is an n x 1 vector of disturbances.

The disturbance μ is taken to be normally distributed with a general diagonal covariance matrix Ω. The diagonal elements allow for heteroskedasticity as a function of exogenous variables, plus a constant term. Z is an n x r matrix of exogenous variables, plus the constant term. The α term is an r x 1 vector of parameters associated with the terms. When $\alpha=0$, it follows that $h=\sigma^2$, the classic homoskedastic situation.

A class of models termed "spatial" follows from this general spatial characterization and are presented in table 3.1. Four more specifications can be obtained by allowing heteroskedasticity of a specific form.

Parameter Values	Name
$\rho=0, \lambda=0, \alpha=0$	classic linear regression model
$\lambda=0, \alpha=0$	mixed-regressive-spatial autoregressive model (SAR)
$\rho=0, \alpha=0$	linear regression with a spatial autoregressive disturbance (SEM)
$\alpha=0$	mixed-regressive-spatial autoregressive model with spatial autoregressive disturbance.

Table 3.1: A Taxonomy of Spatial Models

These spatial models can be used to estimate the HPF incorporating the appropriate assumption about the underlying spatial relationships. The next chapter focuses on how to specify the structure of the spatial weight matrix.

Chapter 4

THE SPATIAL WEIGHT MATRIX

4.1 Overview

The spatial weight matrix embodies the form of the underlying relationship between units of observations and/or their associated error terms. There are two basic types of spatial weight matrices. The first type establishes a relationship based upon shared borders or vertices of lattice or irregular polygon data, often called contiguity-based spatial weight matrices. The second type establishes a relationship based upon the distance between observations, often called distance-based spatial weight matrices. There are many applications of spatial econometric models in public finance where spatial relationships are based on jurisdictions that share borders, such as counties, census tracks and blocks, and states. However, there are few micro-level (land parcels) studies where spatial relationships are based on shared borders. This is due in large part to the lack of polygon micro-level data sets available.

A characteristic of both types of weight matrices is two general parameters that need to be specified before their construction. The first is the spatial extent of the influence. In other words, the size and shape of the area that encompasses the "neighborhood". For a particular county in the U.S., one needs to determine if the neighbors should be restricted to the other counties which border this county, the counties within a 100 mile radius, the counties that border the counties that border this county, or

only the counties to the east. In the case of a house or parcel of land, neighbors could include all those in its sub-division, its neighboring sub-division, or its entire town.

The second parameter that needs to be specified before construction of the weight matrix is the "power" of the influence. In other words, does every member of a neighborhood exert an equal influence, or does each influence the others differently? Suppose the neighborhood is defined as the counties that share a border. Does each of these counties exert the same influence, or would that influence be different, perhaps based upon the amount of border shared? Consider the case of a house or parcel of land where the neighborhood is defined as those other houses or parcels in its sub-division. Would every member of its sub-division exert the same influence?

The selection of a weight matrix has been described as *ad hoc* and/or *a priori*. As an adjective, the term *ad hoc* is defined as "concerned with a particular end or purpose". In most cases, it is not the intention of the model to predict which type of weight matrix (or the parameter values used to build the weight matrix) is correct. The purpose of including a weight matrix is to correct potential problems due to spatial effects, such as inefficient parameter estimates. The term *a priori* is defined as "presupposed by experience; being without examination or analysis; formed or conceived before hand." In most cases, the choice of the weight matrix is made prior to running the model. In other words, the weight matrix is not estimated as part of the model. However, the type of weight matrix chosen and the values of the parameters chosen may be based upon past experience or intuition.

4.2 Contiguity-Based Weight Matrices

The availability of polygon (or lattice) data permits the construction of contiguity-based spatial weight matrices. There are two basic types of contiguity: rook contiguity and bishop contiguity. Rook contiguity exists when two polygons share a common border. Bishop contiguity exists when two polygons share a common vertex (more often referred to in GIS as a node). In addition, queen contiguity is combination of rook and bishop contiguity. Figure 4.1 illustrates rook, bishop, and queen contiguity. Section A would be contiguous to only sections B and C if rook contiguity was imposed. Section A would by contiguous only to section D if bishop contiguity was imposed. Under queen contiguity, section A is contiguous to B, C, and D.

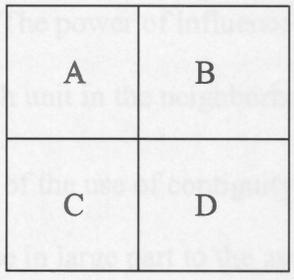


Figure 4.1: A Typical Lattice Structure

A typical specification of the contiguity relationship in the W matrix is given in equation (4.1).

$$w_{ij} = 1, \text{ if i and j are contiguous}$$
$$w_{ij} = 0, \text{ if i and j are not contiguous} \quad (4.1)$$
$$\text{where } w_{ij} \text{ is the i, jth element of W}$$

Another type of contiguity is achieved when a block structure is imposed upon queen contiguity. An example would be contiguity between counties in a particular state but no contiguity between bordering counties which lie in neighboring states (Anselin, 2002; Case, 1991; Case, 1992).

The rook, bishop, and queen contiguity structure define the "neighborhood" parameter suggested previously. The power of influence in either of these general structures is straightforward. Each unit in the neighborhood exerts the same influence.

There are many examples of the use of contiguity-based spatial weight matrices in econometric literature. This is due in large part to the availability of polygon GIS data sets such as states, counties, watersheds, census tracts, and census blocks. The borders of these data sets are well defined and the polygons typically fill the entire landscape.

4.3 Distance-Based Spatial Weight Matrices

Another class of spatial weight matrices is distance-based. These types of spatial weight matrices are based upon the distance between observations. Distance-based spatial weight matrices are widely used in applications where the polygon data does not

exist or is not appropriate. Examples include locations of houses, businesses, sales transactions, airports, parks, etc. The commercial availability of inexpensive Global Positioning Satellite (GPS) handheld units and the lifting of government restrictions on GPS signals have resulted in increased availability of point data. In addition, most statistical packages contain routines to build distance matrices from a data set containing latitude and longitude. A typical specification of the relationship in the W matrix is presented in equation (4.2).

$$w_{ij} = 1/(d_{ij})^{\theta}, \text{ if } i \neq j \text{ and } d_{ij} < m$$
$$w_{ij} = 0, \text{ if } i \neq j \text{ and } d_{ij} \geq m, \text{ or if } i = j. \quad (4.2)$$

The term d_{ij} is the distance between two points i and j within the spatial unit. The point within the spatial unit is typically its geographic centroid. The θ parameter specifies the "power" of the influence described previously. It allows the *decay* effect to be linear if $\theta = 1$ or of higher order for values greater than one. As θ increases, the influence of nearby points becomes greater than points further away.

The parameter m is the "neighborhood" of influence. The choice of a value for m is an empirical question that depends on several factors, including the scale of the data and the extent of the perceived neighborhood. Bell and Bockstael (2000) provide an example of a hedonic model in which the size of the neighborhood was picked based upon the average size of housing developments in an area.

4.4 Cliff and Ord's Weight Matrix

The spatial weight matrices presented in the previous two sections are cases of the more general specification presented by Cliff and Ord (1973). The formal expression of the relationship in Cliff and Ord's spatial weight matrix is presented in equation 4.3.

$$w_{ij} = [d_{ij}]^{-\theta} * [B_{ij}]^{\delta} \quad (4.3)$$

In equation 4.3, d_{ij} = distance between spatial units i and j; B_{ij} = the proportion of the interior boundary of unit i in contact with unit j, and θ and δ are parameters. In the case of counties, equation 4.3 would take into account both the distance between their centers and the length of their common border.

4.5 Spatial Weight Matrices with Estimated Parameters

The contiguity-based, distance-based, and Cliff and Ord weight matrices contain parameters whose values are typically set *a priori*. However, there are examples where parameters of the spatial weight matrix are estimated jointly with the model. Bolduc et al. (1992) present an example of a travel flow model where parameters of the spatial weight matrix are estimated jointly with the other unknown model parameters. LeSage and Pace (2000) introduce a flexible specification for the weight matrix that, when combined with the matrix exponential spatial specification (MESS), leads to a Bayesian model that allows posterior inferences regarding the magnitude and extent of spatial

influence. Lima (1999) adopts a maximum likelihood approach and also a Bayesian Sampling-Importance-Resampling (SIR) procedure to estimate the parameters of a weight matrix.

4.6 Row Standardization of the Spatial Weight Matrix

It is common practice in maximum likelihood (ML) estimation of spatial models to row-standardize the weight matrix. Row-standardization is the process of normalizing each of the weights in a row such that they sum to one. Row-standardization is appealing from both a statistical and computational standpoint. Although row-standardization is useful, it may change the intended "economic" relationship between observations. Consider the following top two rows of a 4 x 4 spatial weight matrix:

Actual Distance			
0	3	6	9
3	0	20	30
.	.	0	.
.	.	.	0

1 / Actual Distance			
0	.33	.17	.11
.33	0	.05	.03
.	.	0	.
.	.	.	0

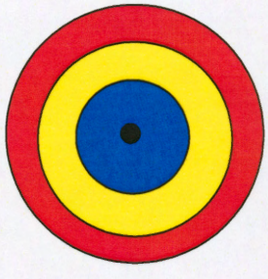
Row-Standardized Distance			
0	.54	.28	.18
.81	0	.12	.07
.	.	0	.
.	.	.	0

Figure 4.2: Comparing the Top Two Rows of Actual, Inverse, and Row-Standardized Inverse Distance Weight

The distance between unit 1 and unit 2 is 3 meters. However, the row-standardized "distance" between them is considerably different. While unit 2 has a potential spatial dependence with unit 1 of .54, unit 1's potential spatial dependence with unit 2 is .81. Consider unit 4's potential influence on unit 1. It is nearly equal to unit 3's potential influence on unit 2. However, unit 3 is more than twice as far away from unit 2 as unit 4 is from unit 1. These two examples show that row-standardizing changes the hypothesized influence. Careful consideration needs to be applied when interpreting results on the parameters of row-standardized weight matrices.

4.7 Bell and Bockstael's Generalized-Banded Spatial Weight Matrix

Bell and Bockstael (2000) proposed a generalized-banded approach to developing a spatial weight matrix. Motivated mainly by the desire to address row-standardization issues, they suggested a flexible spatial weight matrix (contains two separate weight matrices) built with bands that are based on distance. Figure 4.3 shows how the bands



are built.

Figure 4.3: Bell and Bockstael's Bands of 0-200, 200-400, and 400-600 meters

The inner band represents an area of 200 meters around the parcel; the middle

band represents an area between 200 and 400 meters around a parcel; the outer band represents an area between 400 and 600 meters from the parcel. The full weight matrix is row-standardized by row-standardizing each of the individual weight matrices. Parcels that are these distances from the each parcel are given equal weights; otherwise the off-diagonal term is set to 0.

Bell and Bockstael developed a hedonic model of residential parcel prices that included measures of the spatial pattern of land use surrounding each parcel. Their model was estimated using the generalized-moments approach. This approach allows a spatial dependence parameter to be estimated for each weight matrix. Using the flexible form of the spatial weight matrix, they found that eight of their estimated coefficients were significant, compared with five in an inverse-distance based specification of a spatial weight matrix.

Chapter 5

SPATIAL DEPENDENCE AND GEOGRAPHIC BOUNDARIES

5.1 Overview

As mentioned above, the choice of a spatial weight matrix for the error term is a difficult decision to make. The error term captures, among other things, unobservable variables that may be spatial related that are not included in the model. The difficulty presents itself when a researcher asks two important questions: 1) What are the potential unobservable variables? and 2) what is the spatial nature of those variables? An example is air pollution. Levels of air pollution may be unobservable to the researcher. Levels will likely vary based upon wind and pollution sources. Because air pollution knows no "geographic boundaries" per se, the correlation among levels of air pollution at different residential parcel sale locations may be based upon the distance between the respective locations. That is, the level of air pollution at a particular residential parcel is likely more similar to the residential parcels near it than the residential parcels further away.

5.2 Unobservable Quality of Access to Local Parks

The quality of local park access is based upon proximity to the park and the obstructions in a homeowner's path to the park. Local park access is often considered a desirable characteristic when purchasing a house. While it may be simple to measure this distance to the park from any house, it may be more difficult to assess the quality of the access. Consider the map of two sub-divisions in figure 5.1. Separating these sub-divisions is a busy highway that is dangerous to cross on foot or bike. The quality of access to the park by the residents of Park Access Estates is higher than the residents of Pleasant View Estates simply because they don't have to cross a busy highway to get there. If distance to the park is either not included in the hedonic model or is not a good proxy of the quality of park access, then the amenity value of park access will be captured in the error term. Any two of the parcels that border the park are likely to have some potential spatial dependence in their respective error terms resulting from their close proximity to the park. However, a person who is relatively close to the park, but resides across the highway in Pleasant View Estates, would likely have less potential spatial dependence with someone who borders the park, or is of the same distance away from the park, but resides in Park View Estates. In this example, it seems reasonable to restrict the set of influential neighbors to those within a sub-division, as opposed to an "artificial" neighborhood determined by a distance from a particular residential parcel.

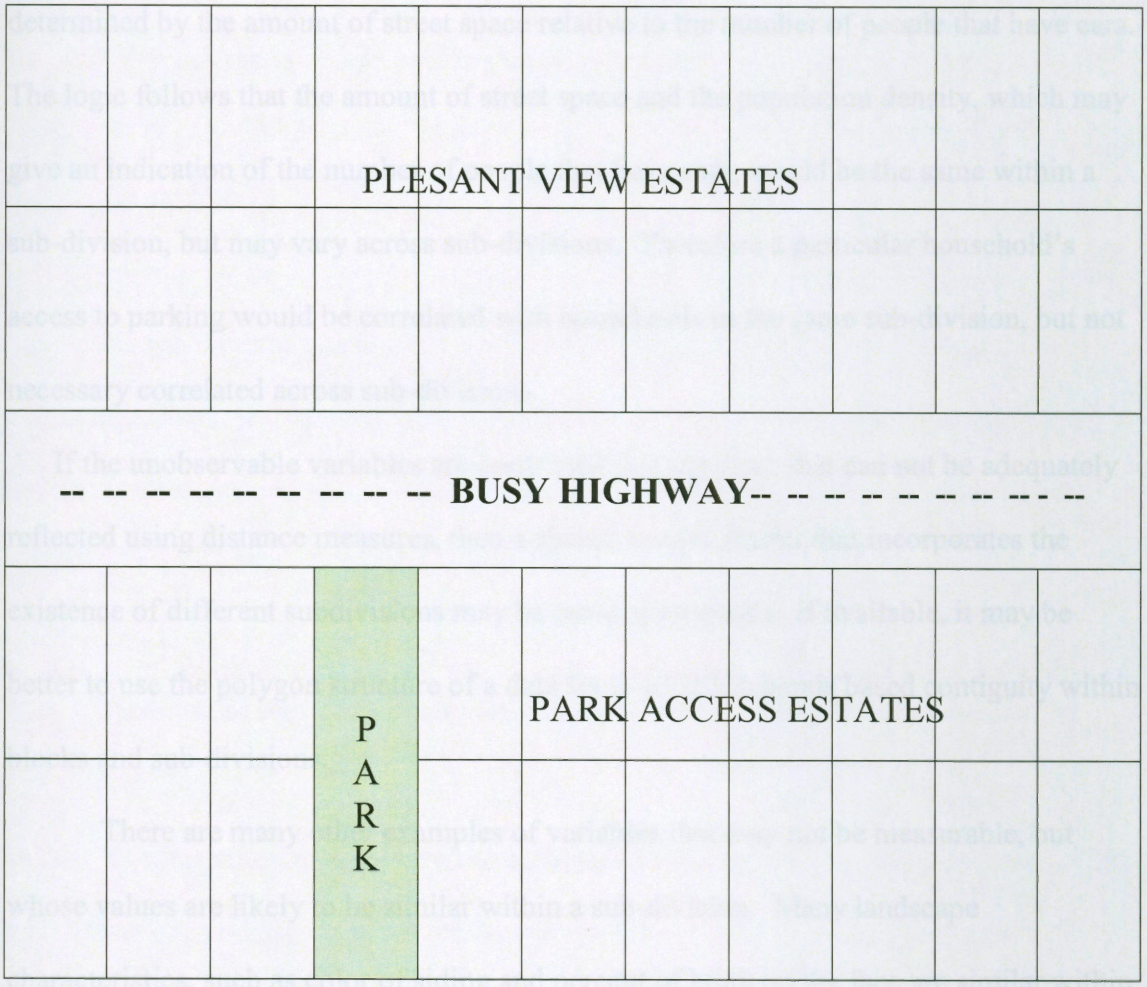


Figure 5.1: Quality of Access to Local Park

5.3 Unobservable Quality of Access to Parking

In an urban setting where population density is high, parking is severely limited, and is often only available on-street. Therefore, it is desirable to live in an area where there is sufficient supply of parking. Without sophisticated GIS analysis, it may be difficult to proxy local parking supply. Therefore, the amenity value of parking may be

present in the error term of a hedonic equation. Local supply of parking is largely determined by the amount of street space relative to the number of people that have cars. The logic follows that the amount of street space and the population density, which may give an indication of the number of people that have cars, would be the same within a sub-division, but may vary across sub-divisions. Therefore a particular household's access to parking would be correlated with households in the same sub-division, but not necessary correlated across sub-divisions.

If the unobservable variables are correlated in a structure that can not be adequately reflected using distance measures, then a spatial weight matrix that incorporates the existence of different subdivisions may be more appropriate. If available, it may be better to use the polygon structure of a data set to establish bands based contiguity within blocks and sub-divisions.

There are many other examples of variables that may not be measurable, but whose values are likely to be similar within a sub-division. Many landscape characteristics, such as color of siding and percent of brick on the face are similar within sub-division, due in part to rules or regulations.[1]

[1] I thank my committee members for this suggestion.

Chapter 6

REVIEW OF LITERATURE ON HEDONIC MODELS

6.1 Overview

Hedonic models have been used to estimate the effect of a wide range of characteristics on the value of a house and/or land. This section of the thesis reviews several studies pertaining to locally desirable and undesirable land uses, often referred to as LDLU's and LULU's, respectively.

6.2 Hedonic Models of Locally Undesirable Land Uses and Disamenities

There is an extensive literature on negative value of undesirable land uses. These LULU's are hypothesized to reduce the value of the residential properties in their vicinity. One LULU that has received attention in the literature is hazardous waste sites. Michaels and Smith (1988) estimated the marginal willingness to pay (MWTP) for removing a hazardous waste site in the Boston area. They estimated the MWTP for removing the closest waste site and found values ranging from $38 to $1799. Another study by Kohlhase (1989) of toxic waste sites in Houston found that the MWTP, evaluated at the means, was $2364 for each mile further away from a toxic waste site.

Noise pollution is a byproduct of some LULU's . Palmquist (1989) estimated the impact of highway noise in three areas of Washington State. Palmquist found that property values decreased from 0.08 to 0.3 percent, depending on the income status of the

affected neighborhood. Espey and Lopez (2000) estimated the relationship between residential property values and airport noise in the Reno-Sparks area of Nevada. They found that houses in the noisier zones have a value of about $2400 less than equivalent houses in the less noisy zones. Furthermore, they found that houses located one mile from the airport would have a market value about $5500 less than the equivalent home located two miles from the airport.

One LULU-related example of environmental degradation that received attention very early in the hedonic literature, and continues to receive attention, is the value of air quality. Ridker and Henning's (1976) investigation of the relationship between air quality and property values is widely regarded as the first example of a hedonic model of housing price. Another more recent example from Chicago is the work of Chattopadhyay (1999). This research focused on both marginal and non-marginal changes in air quality, that is, large changes in air quality. The marginal willingness to pay for an increase of 0.001 units of SO_2 ranged from $878 to $1036, depending on the functional form specification. The non-marginal willingness to pay for a 25% change in SO_2 ranged from $1353 to $2084, depending on functional forms.

Another LULU-related example of environmental degradation that has received attention in the literature is the effects of water quality on housing prices. Leggett and Bockstael (2000) conducted an analysis of water quality's effect on property values along the Chesapeake Bay in Maryland. They found that an increase in water quality of 100 counts of fecal coliform per 100 mL increases property values 2% along the coast of Anne Arundel County. All of their estimates are specific to the location in which they were estimated.

6.3 Hedonic Models of Locally Desirable Land Uses (LDLU's)

Paralleling the study of LULU's and environmental disamenties are studies of locally desirable land uses (LDLU's). One example of a LDLU that received early attention was greenbelts and urban parks (Kitchen and Hendon, 1967; Weicher and Zerbst, 1973; Hammer, Coughlin and Horn, 1974; McMillian, 1974; Correll, Lillydahl and Singell, 1978; Peiser and Schwann, 1993). Tyrävinen (1997) estimated the value of urban forests benefits in Finland. Benefits such as pleasant landscapes, clean air, reduced noise, and recreation benefits were hypothesized to increase the price of apartments near forested areas. The results indicated that a 100-meter increase in distance away from wooded recreation acres decreased apartments' prices by approximately $8 (US). Other similar studies on forestlands have been conducted (Garrod and Willis, 1992a, 1992b, 1992c (in England); Tyrävinen and Meittinen, 2000).

6.4 Hedonic Models of the Value of Open Space in Maryland

Several studies have investigated the value of various forms of open space on the price of residential parcels in the State of Maryland. Geoghegan, Wainger and Bockstael (1997) developed a hedonic model of residential parcel values in a region within a 30-mile radius of Washington, DC (including parts of Montgomery County). Their model went beyond the traditional variables to explain residential parcel values by including variables that summarized the pattern of surrounding land uses. These variables included

diversity and fragmentation indices and portions of open space within various distances. They found that the nature and pattern of the land uses surrounding a residential parcel influenced the price. They also found that their results were highly dependent on the amount of development in the area.

Irwin and Bockstael (2001) estimated a hedonic model of housing sales that included measures of open space. They differentiated their open space variables by categorizing open space into 3 categories: (1) privately owned open space that is developable, including land in crops and pasture and privately owned forests; (2) privately owned open space that is protected from development, including agricultural easements and privately owned conservation areas; and (3) publicly owned open space, including land owned by the federal, state, or county governments. The main issue that is discussed in their paper is the problem of econometrically identifying land use spillovers. They hypothesize that land use spillovers lead to two types of identification problems: endogenous explanatory variables and spatially correlated error terms. They test for these two effects and correct them by using an instrumental variables approach (IV). Their results revealed that the OLS estimation biased the estimated marginal value of various types of open space downwards.

Similar work by Irwin (2002) estimated a hedonic model of housing sales in four Maryland Counties. Irwin's model differed from her previous work with Bockstael by separating open space into a greater number of categories. Open space was first distinguished by whether the land was preserved or developed and then by land use type. Irwin's results revealed that the estimated values for five different types of surrounding open space range from –0.25 to 1.0 percent of the value of the mean property, given a 1

acre change in the amount of surrounding open space. An interesting result was that preserving open space instead of converting it to low-density residential development benefits nearby homeowners by increasing the value of their properties. Geoghegan (2002) also conducted a similar analysis and found that permanent open space increased nearby residential land values over three times as much as an equivalent amount of developable open space in Howard County. Geoghegan, Lynch, and Bucholtz (2002) found in some counties being near preserved open space increased residential values but in other counties there was no effect.

Chapter 7

THE PROPOSED HEDONIC MODEL

7.1 Overview

The first objective of this thesis is to estimate the value of open space using a hedonic model of housing prices. I have suggested a methodological approach somewhat different from past work. I have presented the hypothesis that using geographic delineation at the city block (hereafter referred to as block) and sub-division levels may best capture spatial dependence in the error term, leading to more efficient parameter estimates. I have supported this hypothesis by presenting examples where bands based on block and sub-division may be a more accurate representation of the true structure of spatial dependence, compared to distance based bands. In this chapter of the thesis, I present a hedonic model of residential parcel prices in Montgomery County, Maryland that includes variables for preserved and developable open space near a parcel. The model is tested for spatial dependence in the error term, and if present, incorporates it in the model. The structure of the spatial weight matrices is based on polygon relationships at block and sub-division levels. A comparison is made to spatial weight matrices based on distance bands.

7.2 The Hedonic Model

I propose a hedonic model of housing prices that incorporates spatial dependence in the error term, if needed. The model is:

$$
\begin{aligned}
P &= \alpha + S\beta + L\eta + N\varsigma + T\zeta + D\theta + \epsilon \\
\epsilon &= \lambda_1 W_1 \epsilon + \lambda_2 W_2 \epsilon + \mu \\
\mu &\sim N(0, \sigma^2 I) ,
\end{aligned}
\tag{7.1}
$$

where P is a vector of the most recent sale prices of the parcels (land and improvements); S is a matrix of structural and lot characteristics; L is a matrix of surrounding land uses; N is a matrix of neighborhood characteristics; T is a matrix of sale date dummies; D is a matrix of school district dummies; W_1 and W_2 are spatial weight matrices; α, β, η, ς, ζ, and θ are the corresponding parameter vectors; λ_1 and λ_2 are the spatial autoregressive parameters; and μ is a vector of random error terms.

S is an n x 13 matrix of structural and lot characteristics that are hypothesized to influence the price of the house. These include: construction quality (good and very good, omitting fair and average); construction material (frame and aluminum, omitting all other materials); square foot of structure; age of structure (3 groups, omitting 2 groups); presence of air conditioning; presence of a garage ; number of bathrooms; number of stories; and the lot size. The variables that indicate presence of a particular amenity will be binary.

L is an n x 8 matrix of surrounding land-uses indices. The indices are percents of

four categories of land uses within two buffers of different radii. The four categories of land use include permanently preserved open space, developable open space, low-density residential, and high-density residential. Commercial and water-wetlands-beaches are the omitted variables. Water-wetlands-beaches are omitted because there is very little of this type of land use in Montgomery County. Commercial land use is omitted to avoid perfect multi-collinearity. Permanently preserved land includes agricultural and forest land enrolled in agricultural land preservation programs, golf courses, and state and county parkland. Developable open space include agricultural and forest land that is available for future conversion. Low density residential housing has less than 2 houses per acre. High-density housing has more than 2 houses per acre.

There are two general lengths of buffer radii. The first general length is used to build a buffer that represents a parcel's viewshed, or what a person may be able to see standing on his or her porch. Radii of lengths ⅛ and ¼ mile from the parcel's boundaries were used to construct the buffers. The second general length is used to build a buffer that represents a parcel's neighborhood. Radii of lengths 1 and 1½ miles from the parcel's boundaries are used to construct these buffers. Two different radii for both the viewshed and the neighborhood are included to determine how robust the regression results are to changes in the radii length. Two sizes for the viewshed and two sizes for the neighborhood resulted in four possible *viewshed-neighborhood* combinations. Figure 7.1 provides a map with four possible buffers.

N is an n x 4 matrix of neighborhood characteristics. Montgomery County, Maryland borders Washington, D.C. to the north. Washington, D.C. is the major employment center and proximity to it is hypothesized to be desirable (lower commuting

time, more diverse services). I proxy for the desirability to live near Washington, D.C. by using both road distance to the employment center of DC (estimated at the White House) and road distance to the nearest mass transit train station.[2] The population density and median household income at the block group level from the 2000 Census act as proxies for social-economic conditions at the parcel.

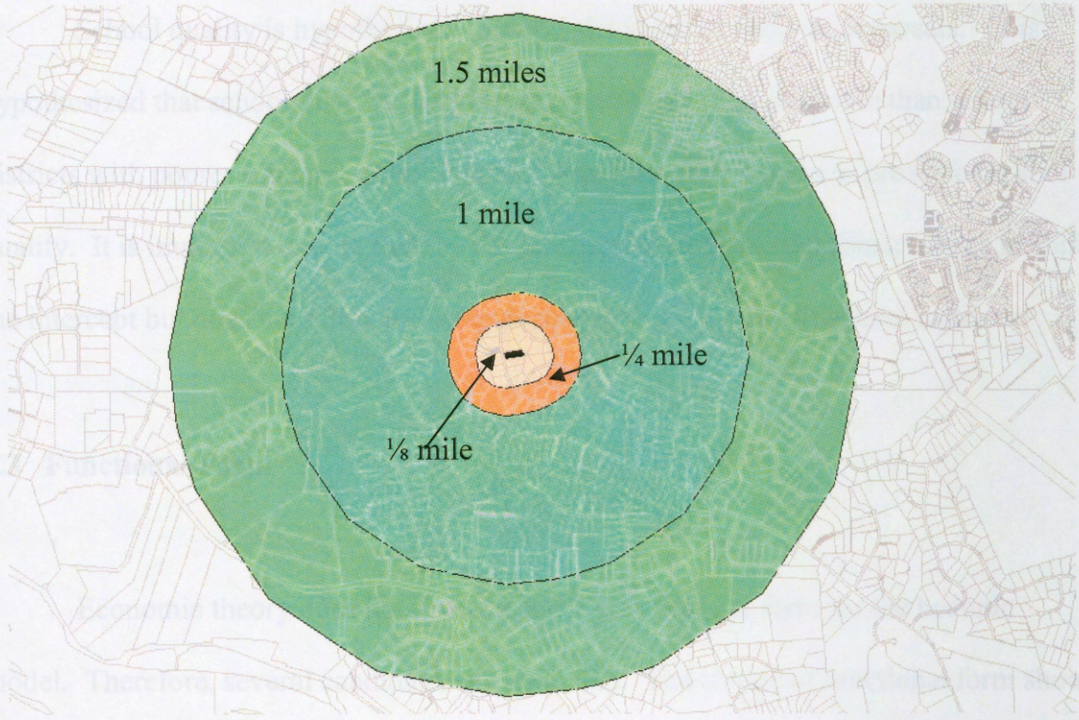


Figure 7.1: Land-use Buffers Sizes: Two View-sheds and Two Neighborhoods

T is an n x 2 matrix of time dummies. Time dummies are included to account for seasonal changes in the market for residential parcels. The residential parcel transactions spanned the time period from October 1, 1996 through September 31, 1998. Evidence

[2] Mass transit train stations were limited to Metro and MARC stations.

has shown that the market for residential parcels fluctuates throughout the year, with higher volume of sales in the late spring, summer, and early fall. The annual market is split into two time periods. The slower period begins on October 1 and ending on March 31. The busier period begins on April 1 and ends on September 30. Therefore, four time periods are included. The two busier time periods are included in the model, as they are expected to have the largest effect.

School quality is hypothesized to affect the value of residential parcels.[3] It is hypothesized that school districts with better schools are more desirable than school districts with poorer schools. School district dummies are included to proxy for school quality. It is important to note that the model specification permits district status to shift the intercept but does not allow the other parameters to vary across school districts.

7.3 Functional Form

Economic theory does not reveal a specific functional form for the hedonic model. Therefore, several candidates are available. The choice of functional form should reflect a researcher's belief. The model presented in equation 7.1 is estimated using log-linear and log-log functional forms.

Another candidate is flexible functional forms such as Box-Cox. However, the many parameters estimated in the Box-Cox transformation reduce the accuracy of any one coefficient. In addition, Box-Cox transformations in the presence of spatial

[3] The statement is based on personal communication with several real estate agents from Long and Foster, as well as one independent property appraiser.

dependence increases the computational difficulty significantly (Pace, Barry, Slawson, Sirmans, 2002).

7.4 The Weight Matrix for the Error Term

The spatial weight matrix in the error term captures spatial dependence among the unobservable variables not included in the model. I use a weight matrix based upon two distinct levels of geographic influence.

The first level of influence is based upon geographic contiguity at the block level. Blocks are sections of parcels that are continuous to each other and typically surrounded by roads on all sides. The first level of influence is meant to capture correlation among unobservable variables that may occur at the very local level. One example may be the presence of a park with unobstructed access. Another example may be an undesirable neighbor, such as a junkyard or a drug dealer.

The second level of influence occurs at the sub-division level. The members of this weight matrix include those who are in a parcel's sub-division, but not in its block. Figure 7.2 provides a local map that illustrates the block (blue area) and non-block/subdivision (red area) neighbors for a residential parcel.

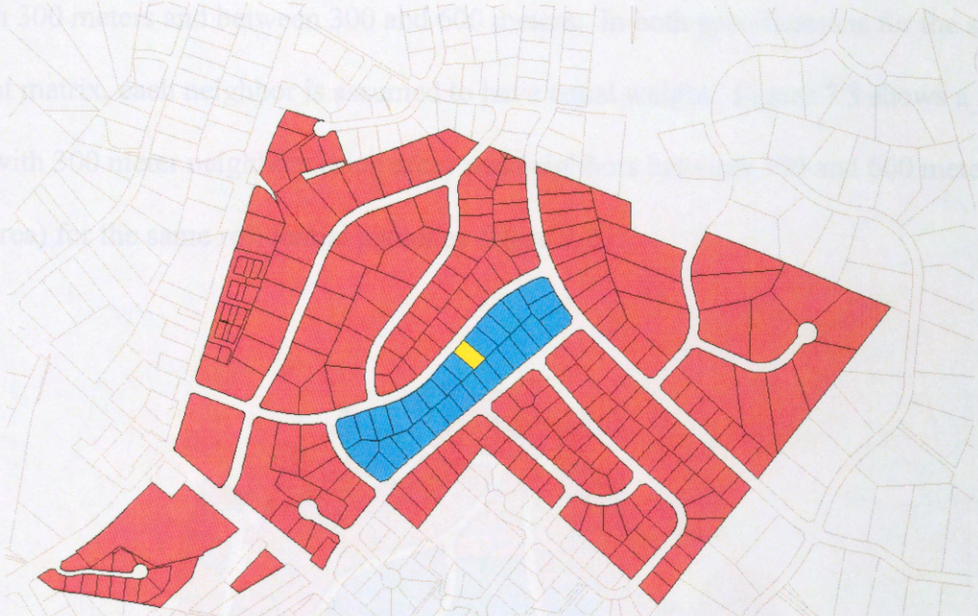


Figure 7.2: Local Map of Block (Blue Area) and Sub-division (Red Area) Neighbors for a Residential Parcel

7.5 Alternative Weight Matrices

This thesis hypothesizes that the true structure of spatial dependence in the error term of a hedonic model is best captured with a flexible form for the spatial weight matrix that allows spatial dependence to vary across portions of a sub-division (i.e., blocks). The hypothesized structure allows dependence within a block and within a subdivision but outside the block. The row-standardized block and non-block sub-division weight matrix will be compared to an alternative specification similar to the one used by Bell and Bockstael. This specification allows dependence among neighbors

within 300 meters and between 300 and 600 meters. In both specifications for the spatial weight matrix, each neighbor is assumed to have equal weight. Figure 7.3 shows a local map with 300 meter neighbors (blue area) and neighbors between 300 and 600 meters (red area) for the same residential parcel as figure 7.2.

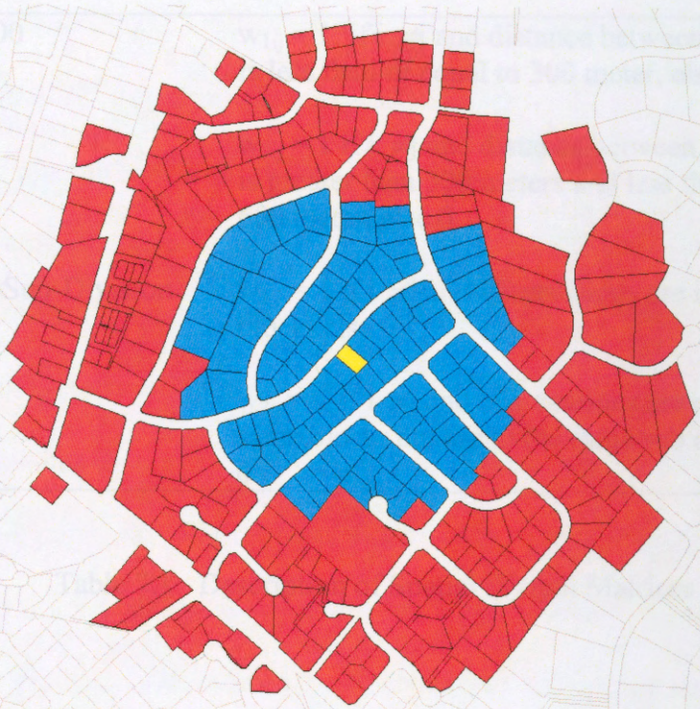


Figure 7.3: Local Map of 300 (Blue Area) and 300-600 (Red Area) Neighbors for a Residential Parcel

The following table summarizes the two different weight matrix specifications that will be used in estimation of the model presented in section 7.1. Recall that each weight matrix specification is made up of two separate weight matrices.

Name	Description
Flexible 300\600	$w_{1ij} = 1$ if i $\neq$j and distance between observations is less than or equal to 300 meter, else 0 $w_{2ij} = 1$ if i $\neq$j and distance between observations is greater than 300 meters and less than or equal to 600 meter, else 0
Flexible Block\Sub-Division	$w_{1ij} = 1$ if i $\neq$j and observations are in the same block, else 0 $w_{2ij} = 1$ if i $\neq$j and observations are not in the same block, but are in the same sub division, else 0

Table 7.1: Description of Spatial Weight Matrices

7.6 Potential Endogeneity of Land Use Indices

Equation (7.1) contains a vector of land use indices as explanatory variables for the price of a residential parcel. Irwin and Bockstael (2001) suggested that these land-use variables may be endogenous to the residential parcel prices as these parcels are usually part of the same real estate market as the parcels in equation (7.1). Therefore, the same economic factors that affect the value of the *i*th residential parcel will also affect the price of the *j*th parcel whose land-use is being used to explain the *i*th residential parcel's value. The probability that the *j*th parcel is in a developed use or is open space may be in part a

function of its own value in a residential use and whether the *i*th parcel has been developed. In essence, unobservable heterogeneity in the error term is correlated with the land use indices. This causes an identification problem. If (7.1) is estimated using OLS, ignoring that variables in L are endogenous, the estimates of η would be biased.

The Hausman test for endogeneity is used to determine if the problem mentioned in the previous section exists. If the land use indices are found to be endogenous, an instrumental variables (IV) approach is used to estimate the model. In the first stage, each of the endogenous land use measures are used as dependent variables in an equation with the explanatory variables being exogenous factors that are not used in (7.1). These factors are assumed to affect the relative costs and benefits of converting the land to a developed use or maintaining it in a non-developed use. To predict land use patterns, the independent variables used are: soil quality variables describing the depth to bedrock; the erosion potential, permeability, slope, and whether the parcel in located in a flood plain; the proximity to nearest transportation node; and whether a sewer system exists or is planned. These predicted values from the first stage are included in the second stage model, which includes the exogenous structural, neighborhood, sale period, and school district variables.[4] The second stage of the model is estimated using Method of Moments.

[4] In cases where the predicted land use index was negative, the value was set to zero.

Chapter 8

METHOD OF MOMENTS ESTIMATION OF SPATIAL MODELS

8.1 Overview

Kelejian and Prucha (1999) developed a generalized-moments (GM) estimator that can be applied to spatial models.[5] GM uses moment conditions implied by the model of interest to form a system of equations to be estimated. Kelejian and Prucha's estimator is based on three moments of the error term ϵ. The key advantage of this estimator is that the calculation of the estimator for even very large data sets is quite straightforward (Bell and Bockstael, 2000). The method requires some matrix multiplication and the calculation of the trace of W'W. However, it does not require the calculation of the determinant or the eigenvalues of W, problems which plague the maximum likelihood (ML) estimating procedure. However, the GM estimator does not permit the calculation of standard errors for the spatial autocorrelation parameter. However, estimates of the spatial autocorrelation parameter(s) are consistent. Therefore, the resulting estimates of the β and σ in the economic model have the large sample properties of the feasible generalized-least squares (FGLS) estimator.

The parameters of the model presented in 7.2 are estimated using GM estimation. GM estimation is used because it is both amenable to using flexible forms for the spatial weight matrices and is computationally feasible for large numbers of observations.

8.2 Generalized-Moments Approach Applied to the Spatial Error Model with One Spatial Weight Matrix

Consider the general spatial error model with one spatial weight matrix in equation 8.1. The GM estimator for the spatial error model is based on three moments of the error term, μ. The following section is based on a summary of the work of Kelejian and Prucha (1999).[6]

$$\begin{aligned} &Y = x\beta + \varepsilon \\ &\varepsilon = \lambda W\varepsilon + \mu \; so \; that \\ &\varepsilon - \lambda W\varepsilon = \mu \, , and \; mulitply \; by \; W, \\ &W\varepsilon - \lambda W^2\varepsilon = W\mu \end{aligned} \tag{8.1}$$

The two representations of the error structure in (8.1) can be written and simplified as follows,

$$\begin{aligned} &Let \; the \; ith \; element \; of \; \varepsilon, W\varepsilon, \; W^2\varepsilon, \mu, and \; W\mu \\ &be \; \varepsilon_i, \dot{\varepsilon}_i, \ddot{\varepsilon}_i, \mu_i, and \; \dot{\mu}_i \; so \; that \\ &\varepsilon_i - \lambda\dot{\varepsilon}_i = \mu_i \; and \\ &\dot{\varepsilon}_i - \lambda\ddot{\varepsilon}_i = \dot{\mu}_i \end{aligned} \tag{8.2}$$

Squaring, summing, and dividing by N gives,

$$\begin{aligned} N^{-1}\textstyle\sum_i \varepsilon_i^2 + \lambda^2 N^{-1}\sum_i \dot{\varepsilon}_i^2 - 2\lambda N^{-1}\sum_i \varepsilon_i\dot{\varepsilon}_i &= N^{-1}\sum_i \mu_i^2 \\ N^{-1}\textstyle\sum_i \dot{\varepsilon}_i^2 + \lambda^2 N^{-1}\sum_i \ddot{\varepsilon}_i^2 - 2\lambda N^{-1}\sum_i \dot{\varepsilon}_i\ddot{\varepsilon} &= N^{-1}\sum_i \dot{\mu}_i^2 \end{aligned} \tag{8.3}$$

---

[5] The phrases "generalized-moments" and "method of moments" are used interchangeably in this thesis and are the same estimation procedure.

[6] This summary was kindly provided to me by Mark Fleming.

Multiplying both representation of the error together and dividing by N gives,

$$N^{-1}\sum_i \varepsilon_i \dot{\varepsilon}_i + \lambda^2 N^{-1}\sum_i \dot{\varepsilon}_i \ddot{\varepsilon}_i - \lambda\left(N^{-1}\sum_i \varepsilon_i \ddot{\varepsilon}_i + N^{-1}\sum_i \dot{\varepsilon}_i^2\right) = N^{-1}\sum_i \mu_i \dot{\mu}_i \qquad (8.4)$$

The three equation system represented in (8.3) and (8.4) includes three moment conditions on the right hand side. These moment conditions can be expressed as

$$\begin{aligned} E[\frac{1}{N}\mu'\mu] &= \sigma^2 \\ E[\frac{1}{N}\mu' W'W\mu] &= \frac{\sigma^2}{N} tr(W'W) \\ E[\frac{1}{N}\mu' W'\mu] &= 0 \end{aligned} \qquad (8.5)$$

The third equality is implied because the diagonal elements of W are always set to zero. The system of equations and moment conditions can be rearranged in matrix form as,

$$G_N\theta - g_N = v_N(\lambda, \sigma^2)$$

where $v_N(\lambda, \sigma^2)$ *is a vector of residuals,* $\theta = [\lambda, \lambda^2, \sigma^2]$,

$$G_N = \begin{bmatrix} \frac{-2}{N}\hat{\varepsilon}'\hat{\dot{\varepsilon}} & \frac{1}{N}\hat{\dot{\varepsilon}}'\hat{\dot{\varepsilon}} & -1 \\ \frac{-2}{N}\hat{\ddot{\varepsilon}}'\hat{\dot{\varepsilon}} & \frac{1}{N}\hat{\ddot{\varepsilon}}'\hat{\ddot{\varepsilon}} & \frac{-1}{N}tr(W'W) \\ \frac{1}{N}(\hat{\varepsilon}'\hat{\ddot{\varepsilon}} + \hat{\ddot{\varepsilon}}'\hat{\varepsilon}) & \frac{1}{N}(\hat{\varepsilon}'\hat{\ddot{\varepsilon}}) & 0 \end{bmatrix}, \; g_N = \begin{bmatrix} \frac{1}{N}\hat{\varepsilon}'\hat{\varepsilon} \\ \frac{1}{N}\hat{\dot{\varepsilon}}'\hat{\dot{\varepsilon}} \\ \frac{1}{N}\hat{\ddot{\varepsilon}}'\hat{\ddot{\varepsilon}} \end{bmatrix} \qquad (8.6)$$

G_N contains the predictors of ϵ, denoted with a "^" symbol. The vector θ contains the parameters to be estimated. The vector g_N is a vector of residuals. The GM estimator is derived from the quadratic in the residuals, $\nu_N(\lambda, \sigma^2)' \nu_N(\lambda, \sigma^2)$.

Kelejian and Prucha (1999) have shown that the GM estimator is both consistent and asymptotically efficient. The OLS residuals are used as predictors of ϵ. Once OLS has been estimated, the system can be solved using non-linear least squares, imposing the necessary additional restriction between the parameters λ and λ^2. The solution for the spatial parameter is both consistent and asymptotically efficient. Once a solution for the spatial parameter is found, estimates for the vector of exogenous variables and model variance can be derived using feasible generalized least squares (FGLS). The feasible and true GLS estimator will be asymptotically equivalent.

8.3 Generalized-Moments Approach Applied to the Spatial Error Model with Two Spatial Weight Matrices

Consider the general spatial error model with two weight matrices in the error term, present in equation 8.7. The GM estimator for the spatial error model is based on six moments of the error term, μ. The following section provides the equations for the case of two weight matrices.[7]

[7] I would like to thank Kathleen Bell for her assistance in completing this section. Any errors are the fault of the author.

$$
\begin{aligned}
&Y = x\beta + \varepsilon \\
&\varepsilon = \lambda_1 W_1 \varepsilon + \lambda_2 W_2 \varepsilon + \mu \; so \; that \\
&\varepsilon - \lambda_1 W_1 \varepsilon - \lambda_2 W_2 \varepsilon = \mu
\end{aligned}
\tag{8.7}
$$

Two additional representations of the error term can be achieved by multiplying the representation in (8.7) (in terms of μ) by W_1 and W_2.

$$
\begin{aligned}
&W_1\varepsilon - \lambda_1 W_1^2 \varepsilon - \lambda_2 W_1 W_2 = W_1 \mu \\
&W_2\varepsilon - \lambda_1 W_2 W_1 \varepsilon - \lambda_2 W_2^2 = W_2 \mu
\end{aligned}
\tag{8.8}
$$

The three representations of the error structure in (8.7) and (8.8) can be written and simplified as follows,

$$
\begin{aligned}
&Let \; the \; ith \; element \; of \; \varepsilon, W_1\varepsilon, \; W_1^2\varepsilon, \; W_2\varepsilon, W_2^2\varepsilon, W_1W_2\varepsilon, W_2W_1\varepsilon, \mu, W_1\mu, and \; W_2\mu \\
&be \; \varepsilon_i, \dot{\varepsilon}_{1i}, \ddot{\varepsilon}_{1i}, \dot{\varepsilon}_{2i}, \ddot{\varepsilon}_{2i}, \dot{\varepsilon}_{1i,2i}, \dot{\varepsilon}_{2i,1i}, \mu_i, \dot{\mu}_{1i}, \; \dot{\mu}_{2i} \; so \; that \\
&\varepsilon_i - \lambda_1 \dot{\varepsilon}_{1i} - \lambda_2 \dot{\varepsilon}_{2i} = \mu_i \; and \\
&\dot{\varepsilon}_{1i} - \lambda_1 \ddot{\varepsilon}_{1i} - \lambda_2 \dot{\varepsilon}_{1i,2i} = \dot{\mu}_{1i} \; and \\
&\dot{\varepsilon}_{2i} - \lambda_1 \dot{\varepsilon}_{2i,1i} - \lambda_2 \ddot{\varepsilon}_{2i} = \dot{\mu}_{2i}
\end{aligned}
\tag{8.9}
$$

Squaring, summing, and dividing by N gives three equations. Multiplying these three equations by each other gives another three equations. The resulting six equation system includes six moment conditions on the right hand side. These moment conditions can be expressed as

$$
\begin{aligned}
&E[\frac{1}{N}\mu'\mu] = \sigma^2 \\
&E[\frac{1}{N}\mu' W_1' W_1 \mu] = \frac{\sigma^2}{N} tr(W_1' W_1) \\
&E[\frac{1}{N}\mu' W_2' W_2 \mu] = \frac{\sigma^2}{N} tr(W_2' W_2) \\
&E[\frac{1}{N}\mu' W_1' W_2 \mu] = \frac{\sigma^2}{N} tr(W_1' W_2) \\
&E[\frac{1}{N}\mu' W_1' \mu] = 0 \\
&E[\frac{1}{N}\mu' W_2' \mu] = 0
\end{aligned}
\tag{8.10}
$$

The fifth and sixth equalities are implied because the diagonal elements of W_1 and W_2 are always set to zero.

The system of equations and moment conditions can be rearranged in matrix. The GM estimator is derived from the quadratic in the residuals, $\nu_N(\lambda_1, \lambda_2, \sigma^2)' \nu_N(\lambda_1, \lambda_2, \sigma^2)$.

Chapter 9

THE DATA SETS

9.1 Overview

Data from several sources is used in this thesis. These data sets include housing sales, parcel boundaries, school district boundaries, land use coverages, soil maps, and sewers.

9.2 Housing Sales Data

The housing sale data consists of approximately 10,500 observations of sales from October 1, 1996 through September 30, 1998. This data set was provided by Dr. Nancy Bockstael, University of Maryland. The sales data is restricted to single family homes.[8] Only the transactions that are classified as "arms length transfers" are included in the data set. Only the transactions that are classified as "total sales" are included in the data set. In addition, observations that do not have a match with a record in the parcel boundary data set are not used (less than 1%).

Additional restrictions are placed on the data set to remove observations that are likely to have "shacks", as opposed to houses, such as those that have a large price and lot size, but a small structure.

[8] Other types of housing structures (condominiums, multiple-family homes, and apartment building) were assessed to be sold in a sufficiently different housing market and thus were not included.

Three additional restrictions are placed on the data set to remove potential outliers. The first restriction limits the data set to observations with a sale price of less than or equal to $1,000,000. The second restriction limits the data set to observations with a structure with a square footage of less than or equal to 6000. The third restriction limits the data set to observations with a lot size of less than or equal to 2 acres. These three restrictions decrease the size of the data set by less than 1%

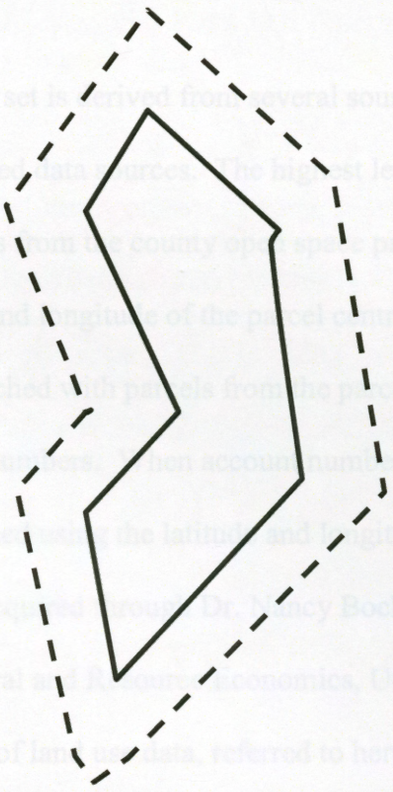
9.3 The Parcel Boundaries Data Set and Buffers

A data set of parcel boundaries in Montgomery County was acquired through Dr. Scott Goetz, Department of Geography, University of Maryland. The source of the data is Montgomery County. The data set contains approximately 270,000 parcels. Each record in the data set contains the account number of the parcel, which is matched with the sales data. After placing the restrictions on the data set mentioned in the previous section, nearly all the remaining observations have a match in the parcel boundaries data set.

Buffers around the parcel boundaries are used for the calculation of land use, soils, and sewer indices. The buffers do not include the parcel itself, so the land use measures do not include own-parcel land use. The ArcView extension *Buffer Theme Builder* was used to construct the buffers. The ArcView extension *Two Themes Analyst*[9]

[9] The Buffer Theme Builder and Two Themes Analyst extensions were written by Kevin O'Mally and are available for download for ESRI's website.

was used to calculate area percents within the buffer for the land use and soils data.



9.1: Buffering a Representative Parcel – Land-use extracted between parcel boundary and buffer edge

Each individual parcel that had sold during the relevant period had the six land-use categories extracted for four buffer lengths. Thus, the data would include how much preserved open space between the parcel edge and the buffer edge. These values were normalized by the area of the buffer and thus represent percent of the land-use within each buffer.

9.4 Land Use Data Sets

The land use data set is derived from several sources. It was built based on a hierarchy of polygon-based data sources. The highest level, referred to hereafter as Level 1, is the preserved parcels from the county open space preservation programs. This data set contains the latitude and longitude of the parcel centroid, as well as the acres preserved. They are matched with parcels from the parcel boundary data set. A match is first made with account numbers. When account numbers are not available or did not agree, the parcel is matched using the latitude and longitude. The records of county preserved parcels were acquired through Dr. Nancy Bockstael and Dr. Lori Lynch, Department of Agricultural and Resource Economics, University of Maryland.

The second level of land use data, referred to hereafter as Level 2, is from the Maryland Department of Natural Resources (MDNR). The data set contains parcel boundaries for the MALPF's agricultural easements, Maryland Environmental Trust lands, private conservation lands, county parks, Federal lands, and MDNR lands (forest and state parks). Level 1 and Level 2 of land use data are merged. Where overlap occurred, Level 1 took precedence.

The last land use data source, referred to hereafter as Level 3, is the 1997 Land Use/Land Cover (LULC) from the State of Maryland Department of Planning. The LULC data set is constructed using satellite imagery and other data sources. It has a resolution of 10 meters. It aggregates land uses into 23 categories. This data set is used only in areas where the Level 1 and Level 2 data sources did not overlap. In other words, Level 1 and Level 2 data sources superceded the Level 3 data set.

Following Geoghegan, Lynch, and Bucholtz (2003), the land use categories are aggregated to the following categories (with LULC codes in parentheses):

Land Use	Corresponding Sources of Data
Preserved open space	Level 1--County preserved parcels Level 2--MALPF agricultural easements, private conservation lands (MET), golf courses and cemeteries(LULC #18), county and Federal parks (DNR)
Developable open space	Level 3--pastureland (LULC #22), cropland (LULC #'s 21,22 23,25), forests (LULU #'s 41, 42, 43, 44, 72, 73), commercial agricultural (LULC #'s 241,242)
Low density residential	Level 3--Properties with 0.2 – 2 dwelling units/acre (LULC #'s 11, 191, 192)
Medium and high density residential	Level 3--Properties with >2 dwelling units/acre (LULC #'s 12, 13)
Commercial	Level 3--Commercial (LULC #s 14, 15 ,16, 17, 18, 80)
Water	Level 3--Water, wetlands, and beaches (LULC #s 50, 60 71)

Table 9.1: Land Use Aggregation Categories and Their Corresponding Sources of Data

Figure 9.2 shows the distribution of the land-use categories across Montgomery County.

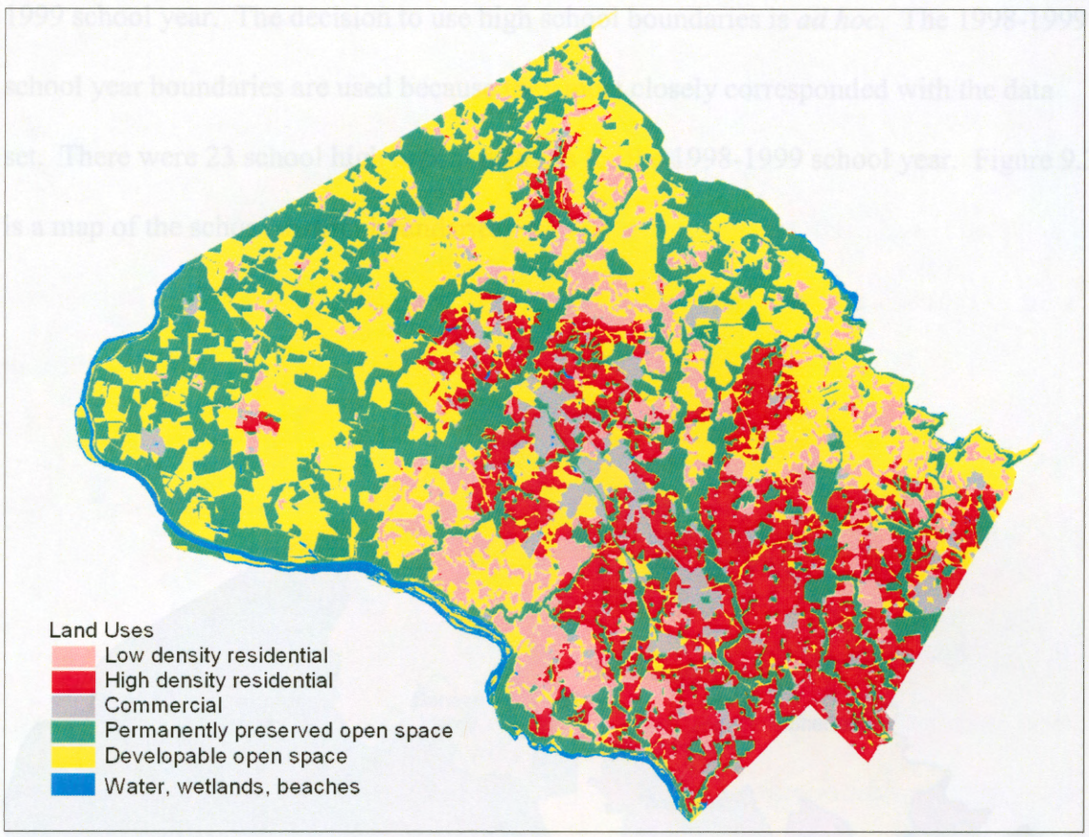


Figure 9.2: Six Aggregated Categories of Land Use, Montgomery County

The sales data spans the time period from October 1996 through September 1998. During that time period, it is assumed that land use patterns did not change significantly.

9.5 School District Boundaries

A data set of school district boundaries was obtained from Jacqueline M. Geoghegan, Department of Economics, Clark University. The data set consists of school

district boundaries for elementary, middle school, and high school. The data set spanned from 1998 to 2003. A decision is made to use the high school boundaries from 1998-1999 school year. The decision to use high school boundaries is *ad hoc*. The 1998-1999 school year boundaries are used because they most closely corresponded with the data set. There were 23 school high school districts for the 1998-1999 school year. Figure 9.3 is a map of the school district boundaries.

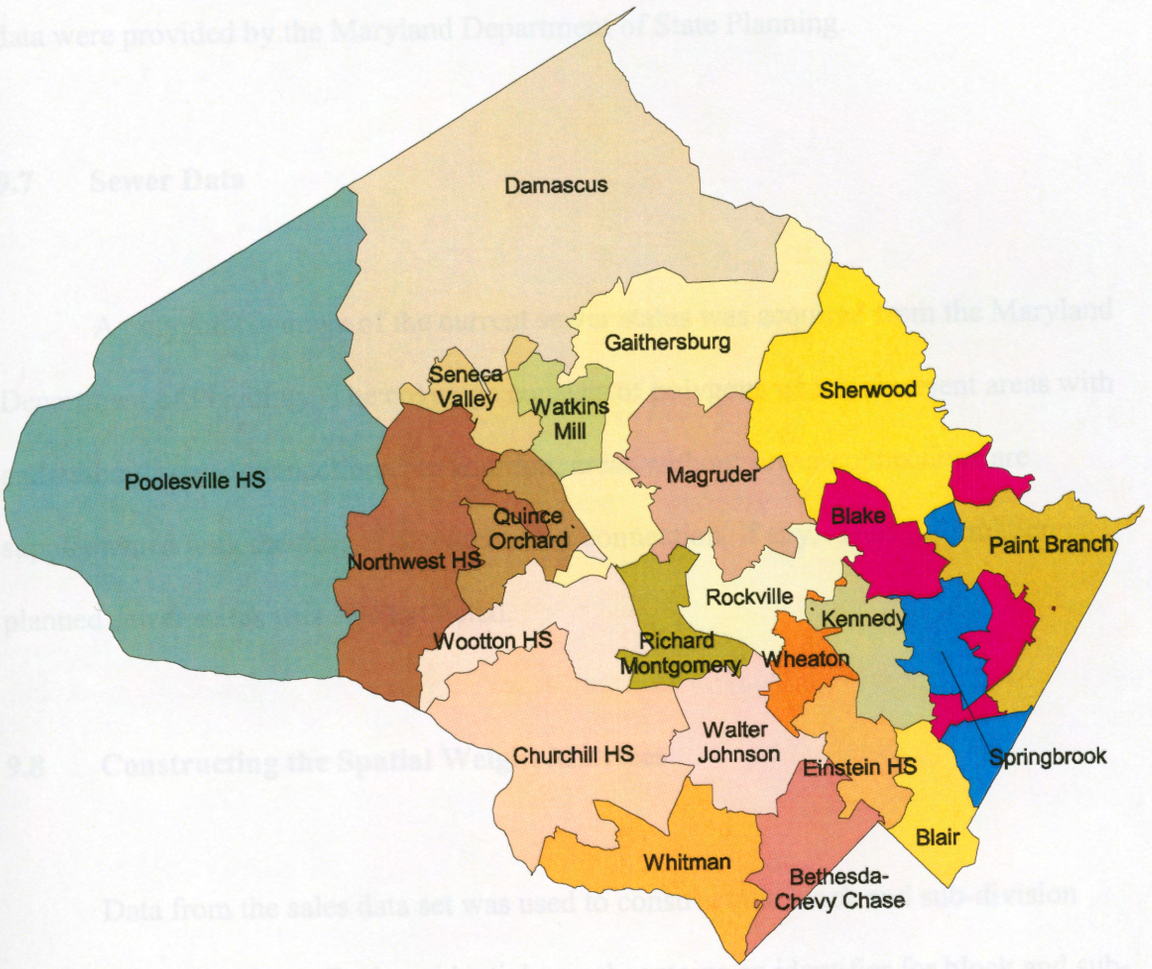


Figure 9.3: Montgomery County High School Districts, 1998-1999

9.6 Soils Data Set

The soils data set was extracted from a polygon coverage with an identifier for each type of soil. Embedded in the identifier is information on the soil's slope, erodibility, permeability, depth to bedrock, and floodplain status. Soils data indices are based on percent of various characteristics within a certain buffer distances of the parcel boundaries. The soils buffers are the same ones used for the land use buffers. The soils data were provided by the Maryland Department of State Planning.

9.7 Sewer Data

A polygon coverage of the current sewer status was acquired from the Maryland Department of Planning. The coverage consists of polygons which represent areas with and without sewer connections. In addition, areas without sewer connections are supplemented with the date of planned sewer connection, if any. Thus information on planned sewer status was also extracted.

9.8 Constructing the Spatial Weight Matrices

Data from the sales data set was used to construct the block and sub-division spatial weight matrices. Each residential parcel contains an identifier for block and sub-division. These values were used in conjunction with simple routines written by the author to construct sparse spatial weight matrices.

The block and sub-division data in the sales data set has some limitations. For some observations, no data for blocks or sub-division values existed and they were not used in this analysis. In addition, some residential parcels did not belong to a sub-division, particularly in rural parts of Montgomery County. In cases where parcels did not belong to sub-divisions, the set of neighbors was restricted to those in the block. In addition, subdivision size varied in the data. Some subdivisions were small and thus each individual parcel would have few "neighbors". Alternatively, some subdivisions were quite large and an additional delineation would have been useful.

9.9 Other Data Sources

A few additional data sources are used to calculate variables used in the hedonic model. Median household income and population density for each Census block in 2000 is derived from polygon and table sources from the Bureau of Census. Distance to the nearest mass transit train stop and distance to Washington, DC are calculated using a road coverage derived from the Tiger line files available from the Bureau of Census. These files were processed by ESRI and are available at the ArcData webpage.[10] The coverage of mass transit train stations was built using McNally's Atlas and Tiger line files.[11]

[10] Data available at http://arcdata.esri.com/data/tiger2000/tiger_download.cfm

[11] The coverage was built by the author and is freely available upon request. It contains Metro and MARC stations for Montgomery County and most of Price Georges County and the Northwest section of D.C.

Chapter 10

EXPLORATORY DATA ANALYSIS

10.1 Overview

Exploratory Data Analysis (EDA) is defined as:

> "an approach or a philosophy for data analysis that employs a variety of techniques (mostly graphical) to: 1) maximize insight into a data set; 2) uncover underlying structure; 3) extract important variables; 4) detect outliers and anomalies; 5) test underlying assumptions; 6) develop parsimonious models; and 7) determine optimal factor settings." (NIST/SEMATECH, 2003, pg. 1)

All of the data used in the research project comes from human sources. In particular, the housing structure data from Maryland Property View contains some data that was entered by humans and is therefore susceptible to errors. This chapter of the thesis presents the results of EDA on each of the main sets of data by looking at summary statistics and distribution plots. Particular attention is paid to point numbers 2, 4, and 5 of the definition of EDA.

10.2 Housing Structure and Neighborhood Variables

The first step in EDA is to look at summary statistics for the variables. After an investigation of an initial set of summary statistics, three sets of anomalies were discovered. Some observations had values for **Price** that seemed rather low, relative to

their structural attributes. An investigation of these observations suggests that there are missing zeros in the prices. A house that has a **Price** of $250,000 may have been miscoded at $25,000. Second, a few observations have values for **Lot Size** that are very large relative to the size of the structure. Third, a few observations have structures that are very small relative to the **Lot Size** and **Price**. The conclusion is drawn that the first set of anomalies are likely miscoded and the second and third sets of anomalies are likely land sales. These observations are dropped from the analysis. Table 10.1 contains the summary statistics after removal of the aforementioned observations.

Variable Name	Mean	Std. Dev.	Min	Max	Skew-ness	Kurt-osis
Price	$272,521	$139,456	$30,000	$1,000,000	1.57	3.07
Structural Variables						
Age	29.44	22.24	1.00	197.00	0.47	0.19
Square footage	2,209	1,054	520	5,991	0.94	0.47
Attic	0.03	0.18	0.00	1.00		
Garage	0.69	0.46	0.00	1.00		
Basement	0.92	0.27	0.00	1.00		
Number of bathrooms	2.48	0.93	1.00	7.00		
Number of stories	1.68	0.44	1.00	3.00	-0.74	-1.20
Central air	0.84	0.37	0.00	1.00		
Lot size (acres)	0.32	0.30	0.04	2.00	3.51	14.25
Fair construction quality	0.03	0.17	0.00	1.00		
Avg. construction quality	0.52	0.50	0.00	1.00		
Good construction quality	0.36	0.48	0.00	1.00		
Very good construction quality	0.09	0.29	0.00	1.00		
Aluminum	0.18	0.38	0.00	1.00		
Wood Frame	0.38	0.49	0.00	1.00		
Other Bldg. Material	0.44	0.50	0.00	1.00		
Neighborhood Variables						
Median household income (@ Cen. Blk.)	$96,711	$34,727	$7,371	$200,001	0.75	0.34
Population density (persons per square mile @ Cen. Blk)	4,361	2,971	35	31,558	1.54	6.07
Distance to Washington, DC (in kilometers from parcel)	23.9	10.2	8.2	56.5	0.53	-0.62
Distance to train station in kilometers from parcel)	4.98	3.50	.10	20.72	1.01	0.75

Table 10.1: Summary Statistics for Structural and Neighborhood Variables

The second step in EDA is to look at various distribution plots of the continuous variables. A histogram, illustrated with a truncated normal and lognormal density curves, is generated for each of the continuous variables. These histograms are presented in Figure 10.1 – 10.8. The y-axis is Percent of the observations and the x-axis is the

Variable in question. The normal density curve is the solid line; the log-normal density curve is the dashed line.

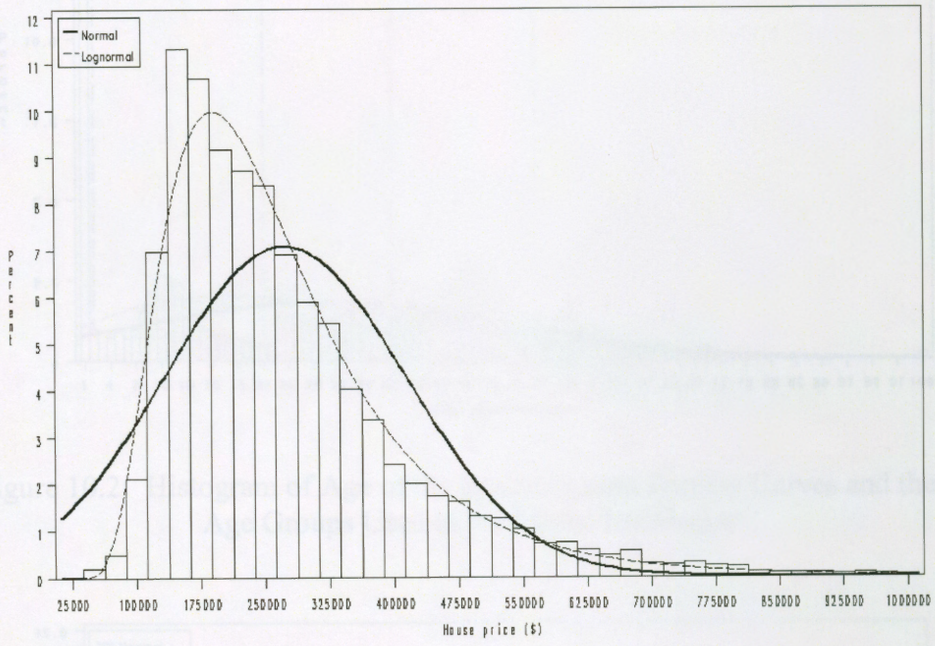


Figure 10.1: Histogram of Parcel and Housing Price with Density Curves

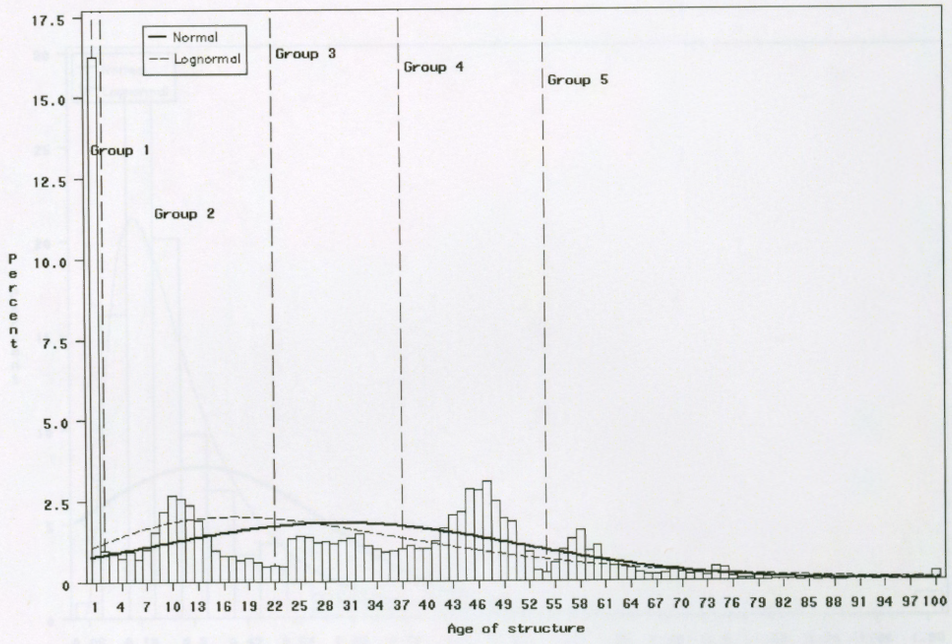


Figure 10.2: Histogram of Age of the Structure with Density Curves and the 5 Age Groups Used in the Model Estimation

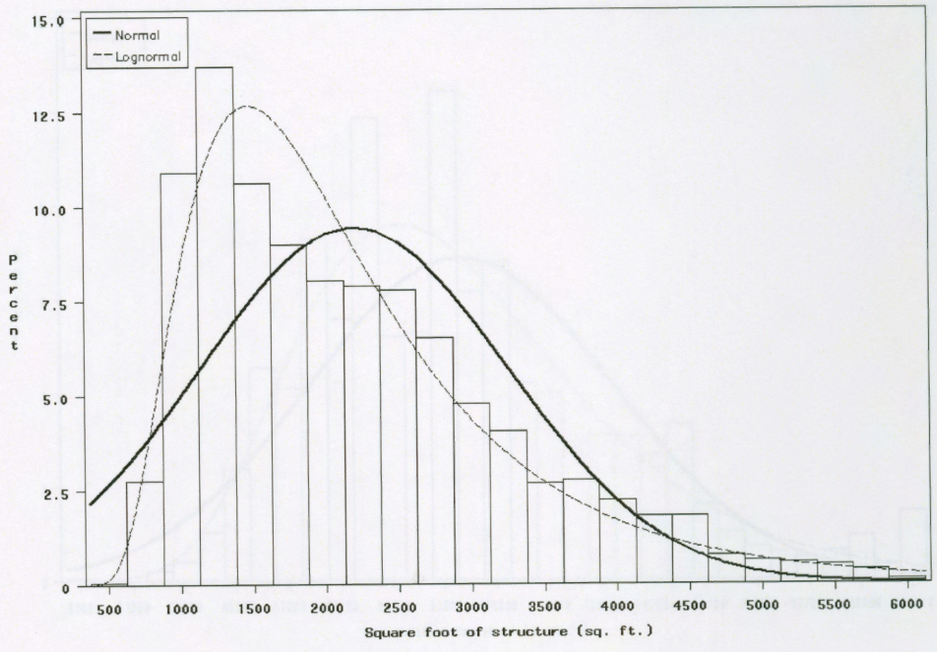


Figure 10.3: Histogram of the Square Foot of Structure with Density Curves

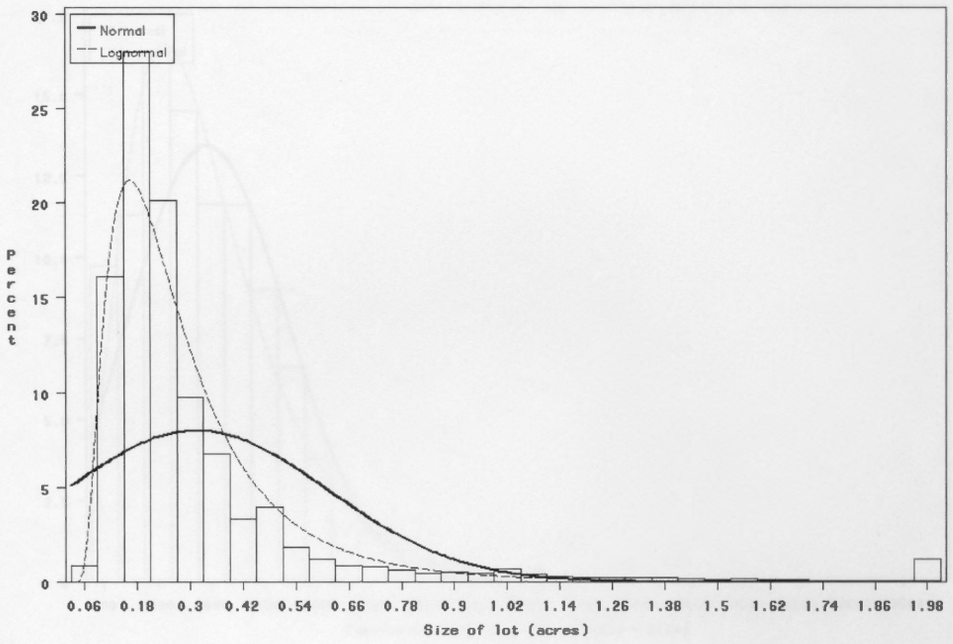


Figure 10.4: Histogram of Size of the Lot (Acres) with Density Curves

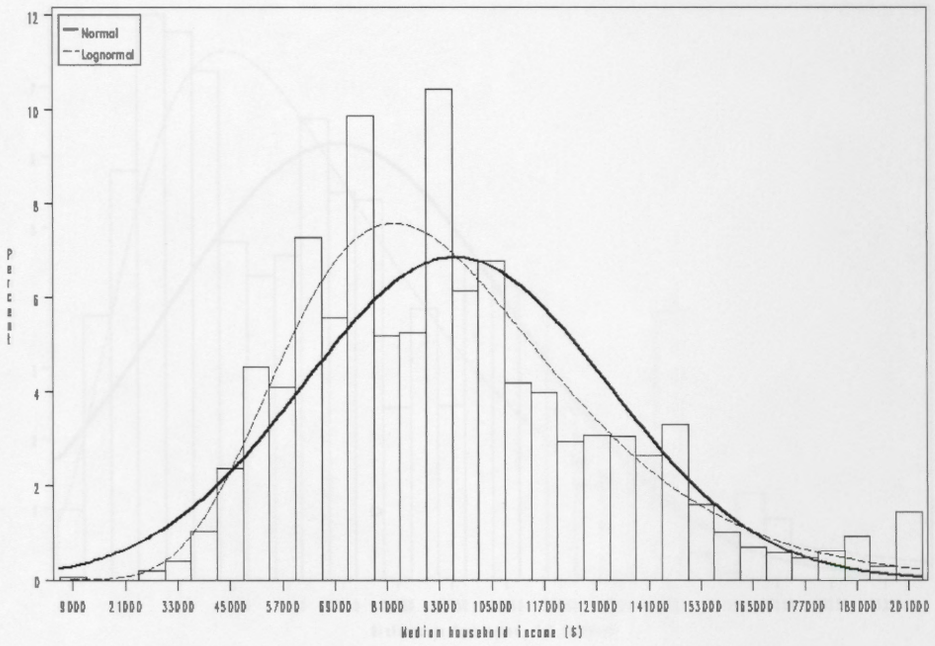


Figure 10.5: Histogram of Median Household Income ($) with Density Curves

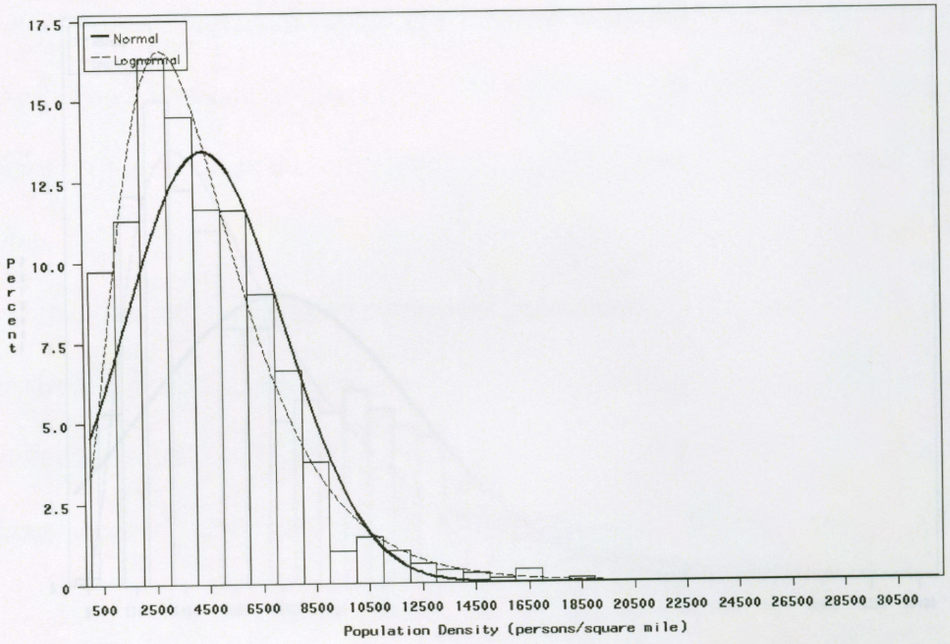


Figure 10.6: Histogram of Population Density (Persons/Square Mile) with Density Curves

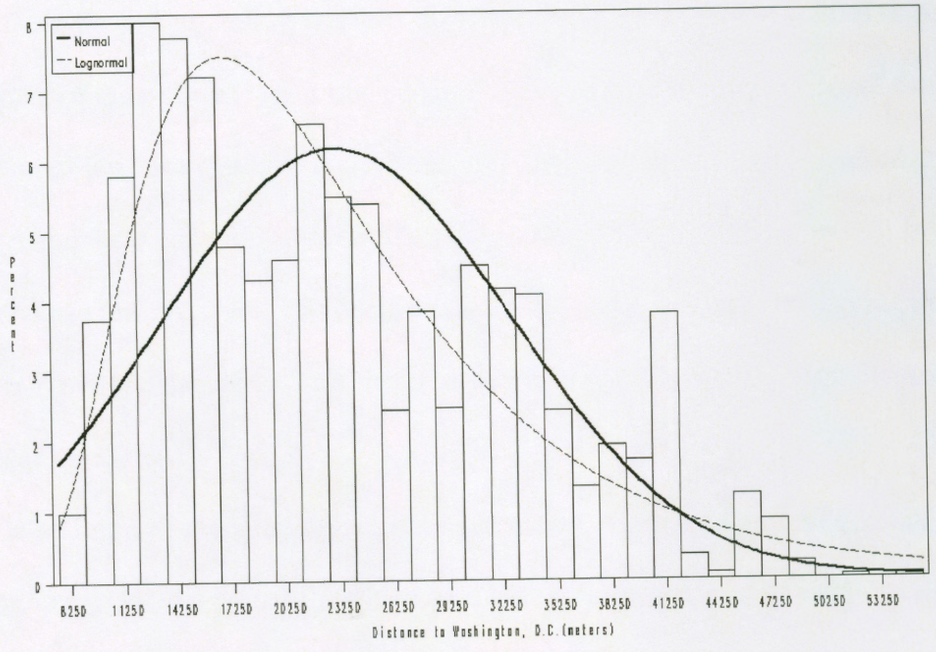


Figure 10.7: Histogram of Distance (Meters) to Washington, D.C. with Density Curves

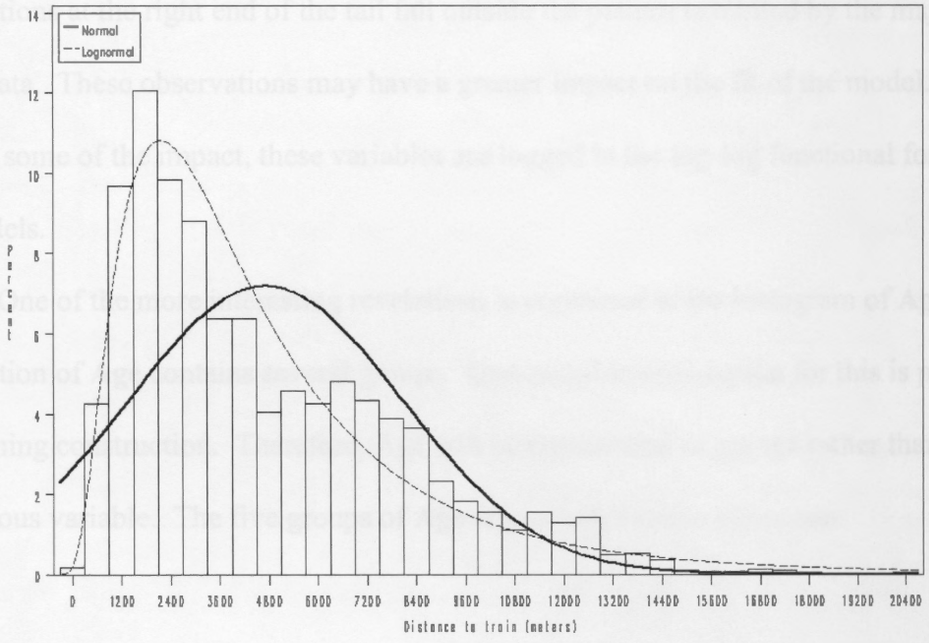


Figure 10.8: Histogram of Distance to Train Station (meters) with Density Curves

The histograms for several of the variables indicate right-skewness. The observations at the right end of the tail fall outside the pattern exhibited by the majority of the data. These observations may have a greater impact on the fit of the model. To remove some of the impact, these variables are logged in the log-log functional form of the models.

One of the more interesting revelations is contained in the histogram of **Age**. The distribution of **Age** contains several groups. One possible explanation for this is periods of booming construction. Therefore, **Age** will be represented as groups rather than as a continuous variable. The five groups of **Age** are included in the histogram.

10.3 Land Use Indices

The type and amount of land use surrounding a parcel is hypothesized to be a determinant of the price of residential parcels. Land use indices are developed based on the percent of land use inside of a viewshed and a neighborhood. The viewshed and neighborhood were constructed for buffers of a certain radii around the parcel. In order to use land use indices, a few key decisions are made. The land uses are aggregated into a small number of categories. This decision was made to best elicit the importance of the open space variables.

In addition to choosing aggregation categories, a determination of the size of the area around a particular parcel for which land use is measured must be made. It was previously suggested that land use should be measured at two distances. The first distance reflects a parcel's "view shed", that is, what an owner may be able to see if

standing on his or her porch. The second distance reflects a parcel's "neighborhood". The choice of the values for either of these distances is arbitrary, but should reflect a researcher's beliefs about what most people can see from their porch and what most people consider their neighborhood. The choice was made to use two distances to represent a viewshed and two distances to represent a neighborhood. This choice was based on intuition and past research. This results in four *viewshed-neighborhood* combinations.

EDA was useful in investigation how the final model results may be sensitive to a change in the viewshed-neighborhood combinations. Table 11.2 presents summary statistics for each of the land use categories at each of the four buffer distances.

	Mean	Std. Dev.
Permanently preserved open space		
⅛ mile viewshed	8.60%	15.84%
¼ mile viewshed	12.00%	14.98%
1 mile neighborhood	18.43%	9.88%
1½ mile neighborhood	19.44%	7.90%
Developable open space		
⅛ mile viewshed	8.33%	16.04%
¼ mile viewshed	10.62%	16.32%
1 mile neighborhood	13.80%	15.11%
1½ mile neighborhood	14.65%	14.76%
High/Medium density residential		
⅛ mile viewshed	70.06%	29.98%
¼ mile viewshed	61.93%	26.67%
1 mile neighborhood	47.63%	20.78%
1½ mile neighborhood	44.64%	19.44%
Low density residential		
⅛ mile viewshed	8.68%	20.96%
¼ mile viewshed	8.75%	17.38%
1 mile neighborhood	9.44%	11.56%
1½ mile neighborhood	10.00%	10.68%
Commercial		
⅛ mile viewshed	4.27%	9.33%
¼ mile viewshed	6.54%	9.41%
1 mile neighborhood	10.26%	7.37%
1½ mile neighborhood	10.75%	6.79%
Water		
⅛ mile viewshed	0.07%	0.94%
¼ mile viewshed	0.15%	1.12%
1 mile neighborhood	0.44%	1.83%
1½ mile neighborhood	0.52%	1.76%

Table 10.2: Summary Statistics for Land Use Indices within Two Viewsheds (1/8 mile and ¼ miles) and Two Neighborhoods (1 mile and 1 ½ mile)

The decision of which distances should be used to approximate a view shed and a neighborhood is somewhat arbitrary, relying on intuition and past research. Table 10.2 provides an idea of how land use percents change as the radius used to build the buffer increases. Table 10.3 is an extension of table 10.2, and focuses specifically on the two types of open space. Table 10.3 contains additional radius sizes that were not used in the final models.

	Mean	Std. Dev.	Skewness	Kurtosis
Permanently preserved open space				
⅛ mile	8.60%	15.84%	2.62	8.33
¼ mile	12.00%	14.98%	1.80	3.80
⅜ mile	14.19%	13.94%	1.31	1.43
½ mile	15.62%	12.91%	1.06	0.80
¾ mile	17.40%	11.30%	0.99	1.14
1 mile	18.43%	9.88%	0.90	1.15
1¼ mile	18.96%	8.73%	0.82	1.17
1½ mile	19.44%	7.90%	0.84	1.33
1¾ mile	19.78%	7.35%	0.90	1.63
2 miles	20.05%	6.97%	0.97	2.20
Developable open space				
⅛ mile	8.33%	16.04%	2.74	8.69
¼ mile	10.62%	16.32%	2.18	5.20
⅜ mile	11.68%	16.11%	1.92	3.88
½ mile	12.38%	15.83%	1.76	3.08
¾ mile	13.33%	15.54%	1.55	2.05
1 mile	13.80%	15.11%	1.42	1.45
1¼ mile	14.23%	14.87%	1.30	1.00
1½ mile	14.65%	14.76%	1.21	0.68
1¾ mile	14.96%	14.72%	1.14	0.42
2 miles	15.27%	14.76%	1.08	0.15

Table 10.3: Summary Statistics for Permanently Preserved and Developable Open Space Indices within 10 Different Buffer Radii

Tables 10.2 and 10.3 provide some interesting results as they relate to the choice of buffer distances. The most striking result is the difference between the means for ⅛ mile and ¼ mile buffers for both types of open space. For permanently preserved open space, the mean amount in the ¼ mile viewshed is almost 50 percent more than the mean in the ⅛ mile viewshed. This difference provides justification for estimating several models using different viewshed-neighborhood combinations.

10.4 The Spatial Weight Matrices

There are two different specification for the spatial weight matrix used to estimate the hedonic model in chapter 7. Each of these specifications of the spatial weight matrix contains two separate spatial weight matrices, for a total of four different individual spatial weight matrices. Table 10.4 provides statistics on the number of neighbors in each spatial weight matrix. The results from table 10.4 shows that the block and sub-division structures use less neighbors, on average, than distance bands of 300 and 600 meters.

	Number of neighbors			
Name	Mean	Std. Dev.	Min	Max
CONTIG 300	16.12	14.12	1	115
CONTIG 300-600*	27.91	18.27	1	115
CONTIG BLK	5.52	6.42	1	40
CONTIG BLK-SUB**	30.53	28.41	1	146

*Number of neighbors between 300 and 600 meters

**Number of neighbors outside of block, but within sub-division

Table 10.4: Summary of the Number of Neighbors in each Spatial Weight Matrix

Chapter 11

ENDOGENEITY AND SPATIAL DEPENDENCE TEST RESULTS

11.1 Tests for Endogeneity

Section 7.6 discussed potential endogeneity in the measures of land use. To address potential endogeneity, an instrumental variables approach to estimating the land use indices is used. The instruments include soils data, the presence of a sewer, and the distance to the nearest road. In the first stage, instruments are used to predict land use indices. The results of the Hausman test are presented in table 11.1. The table shows in which models a land use measure was found to be endogenous. Endogeneity in the viewshed land use index measures is denoted by V while endogeneity in the neighborhood land use index is denoted by N. In the second stage, the full model is estimated with the predicted land use indices substituted for the actual land used indices if endogeneity exists. The Hausman test for endogeneity is used to determine if endogeneity exists.

The most striking result from the Hausman test is the considerable difference in the instances of endogeneity between models that uses different sizes for the viewshed. The land use indices measured in the larger size viewshed (¼ mile radius buffer) were much more often endogenous than the same indices measured in the smaller sized viewshed (⅛ mile radius buffer). In addition, the log-linear functional form had more

instances of endogeneity. However, the differences between the occurrences of endogeneity in the two functional forms (log-linear and log-log) are minor.

	Log-Linear Functional Form				Log-Log Functional Form			
Viewshed/ Neighborhood combination*	⅛ mile 1 Mile	⅛ mile 1 ½ Mile	¼ mile 1 Mile	¼ mile 1 ½ Mile	⅛ mile 1 Mile	⅛ mile - 1 ½ Mile	¼ mile 1 Mile	¼ mile - 1 ½ Mile
Permanently preserved open space			**V**	**V**				
Developable open space	**N**		**V,N**	**V**			**V**	
Low density residential	**N**		**V,N**	**V**		**V**	**V**	
High density residential			**V,N**	**V**	**N**		**V,N**	

* V denotes endogeneity of the land use index in the viewshed, N denotes endogeneity of land use index in the neighborhood

Table 11.1: Results of Hausman Endogeneity Test for Land Use Indices

The substantial number of instances of endogeneity in the viewshed, when measured at ¼ mile, is a curious result. There are interesting differences that lie in the performance of the instruments in the two sizes for the viewshed. The R-squared values in the first stage model for each of the land use indices were higher for the ¼ mile viewshed when compared to the ⅛ mile viewshed. Furthermore, all the R-squared values for the ⅛ mile viewshed were less than 35% suggesting that 65% of the variation in the land-use instances was unexplained. The lack of fit in the ⅛ mile buffer may be the reason for the lack of endogeneity in the ⅛ mile land use indices.

Another interesting result from the first stage regressions is the difference in model fit between the different types of land use indices. The model for permanently

preserved open space had the poorest fit while the model for developable open space had the best fit. Previous work by Lynch and Lovell (2003) suggests that the decision to enroll one's farm in a preservation program is partially based on family characteristics. The available data only contained physical characteristics. This suggests that the instruments chosen may not be appropriate for certain types of land uses.

The results from the tests for endogeneity and the first stage models certainly raise more questions than answers. Are the instruments good? Should endogeneity exist at smaller levels if it is found to exist at large levels? Would the land-use models perform better if they were run simultaneously? While the independent variables used to predict land-use have been used before (Bockstael and Irwin, Irwin 2002, Geoghegan, Lynch and Bucholtz 2003), it could be they are not truly exogenous to the price of residential structures and thus an identification problem plagues the models. Can additional variables be found to deal with this identification problem that could be independent of residential prices? More work on identifying appropriate models and instruments are needed to ensure informative results. The results from this section also suggest that researchers should estimate models with different viewshed sizes when evaluating land use around a residential parcel. The questions raised provide direction for future research.

11.2 Tests for Spatial Dependence

It is hypothesized in chapters 2, 3, and 7 that spatial dependence in the error term exists in hedonic models of housing price. There are various tests for spatial dependence

that can be carried out using the residuals from OLS estimation. Due to the number of observations in the data set, the Lagrange Multiplier error (LM Error) and Likelihood Ratio test statistics are the best choices.

Tests for spatial dependence were based upon four separate single weight matrices where neighbors were specified as observations within 300 meters, observations within 600 meters, observations within the block, and observations within the subdivision.[12] The log-log model with ⅛ and 1 mile viewshed-neighborhood combination is used to calculate the test statistics. Table 11.1 summarizes the weight matrix used and the results of the test statistics.[13] The test statistics suggest that spatial autocorrelation in the error terms is present. While the test statistic values change with the specification of the weight matrix, the overall result does not.

Weight Matrix	LM Error value	Marginal Probability	Likelihood Ratio	Marginal Probability
CONTIG 300	1117	0	769	0
CONTIG 600	1627	0	787	0
CONTIG BLK	507	0	467	0
CONTIG SUB	1379	0	772	0

Table 11.2: Spatial Autocorrelation Test Statistic Results for Single Weight Matrices

[12] Tests for spatial dependence were not based on the hypothesized specification for the spatial weight matrices because of computational issues, which are the reason the model is estimated using Method of Moments

[13] James P. LeSage's Econometrics Toolbox for MATLAB was used to calculate test statistics. This package can be downloaded at http://www.spatial-econometrics.com/.

11.3 Mapping of Errors

Another technique for assessing spatial dependence is to plot the estimated error terms on a map and look for patterns. The large number of observations in the data set makes it difficult to see a clear pattern. However, at higher levels of resolution, patterns emerge. Figures 11.1 and 11.2 are maps of two areas in Montgomery County which exhibit a clear pattern of the values of the residuals. The residuals denoted by red circles are greater than 1 positive standard deviation from the mean and the residuals denoted by blue triangles are less than negative one standard deviation from the mean. Each of the maps shows clustering of circles and triangles in various places. These clusters tend to be within blocks, which supports the hypothesis that spatial dependence may be best captured by assuming relationships between blocks and subdivision neighbors.

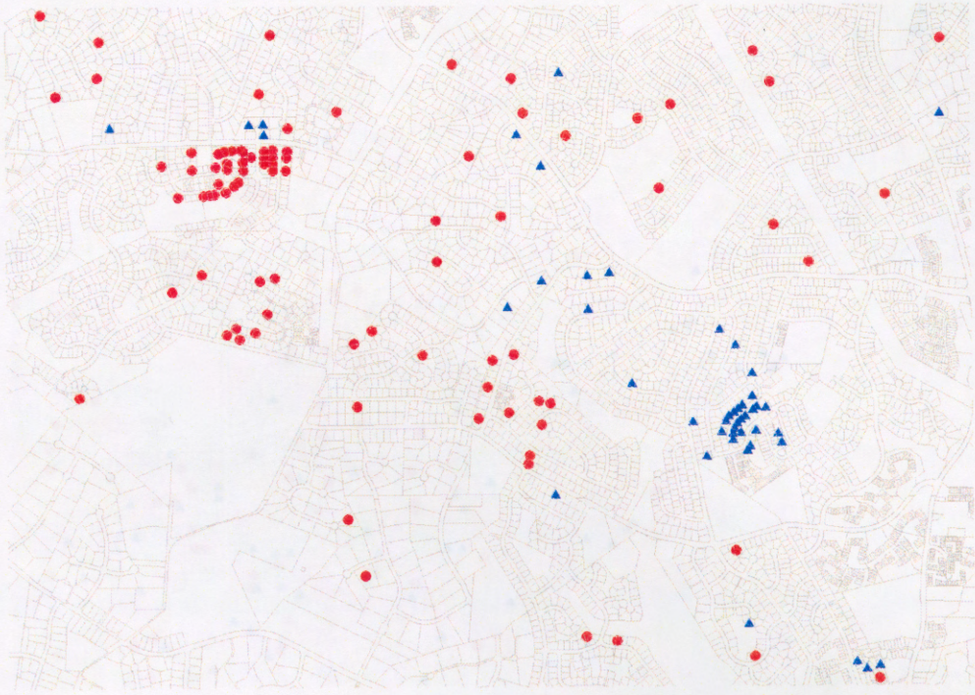


Figure 11.1: Map of OLS Residuals--Germantown Region of Montgomery County: Red Circles are Positive and Blue Triangles are Negative Residuals

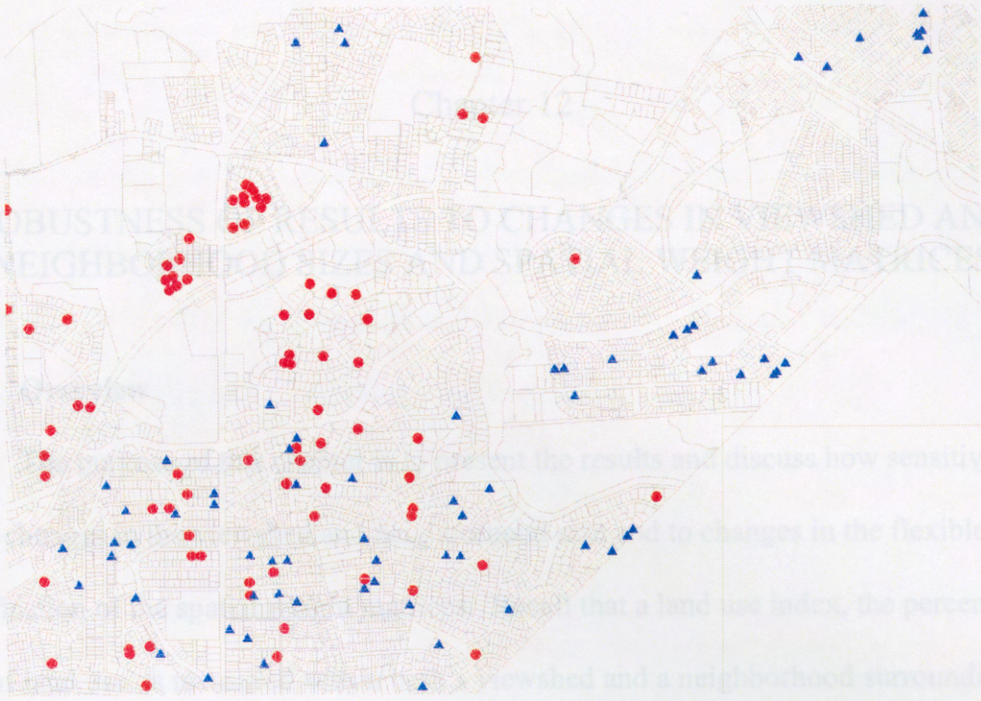


Figure 11.2: Map of OLS Residuals--Silver Spring Region of Montgomery County: Red Circles are Positive and Blue Triangles are Negative Residuals

Chapter 12

ROBUSTNESS OF RESULTS TO CHANGES IN VIEWSHED AND NEIGHBORHOOD SIZES AND SPATIAL WEIGHT MATRICES

12.1 Overview

The purpose of this chapter is to present the results and discuss how sensitive they are to changes in the viewshed and neighborhood size and to changes in the flexible specification of the spatial weight matrices. Recall that a land use index, the percent of a type of land use, is measured within both a viewshed and a neighborhood surrounding a parcel. This thesis attempts to measure the sensitivity to changes in the size of the viewshed and neighborhood by specifying two different sizes for both the viewshed and neighborhood. This results in four viewshed-neighborhood combinations.

The first section of this chapter focuses on the estimated parameters for the structural, neighborhood, and time variables. The second section of this chapter focuses on the robustness of the estimated parameters form land use indices to changes in the spatial weight matrices. The third section focuses on the robustness of the estimated parameters for the land use indices to changes in the viewshed and neighborhood size. The interpretation of the land use indices will be made in Chapter 13.

Reference are made to "qualitative difference" when comparing two (or more) parameter estimates. This means that one parameter estimate is significantly different from zero (at the 95% level) while the other parameter estimate(s) aree not significantly different from zero. References are also made to "quantitative difference" when comparing two or more parameter estimates. This means that a t-test is conducted and

parameter estimates are found to be significantly different from each other at the 95% level.

The model presented in section 7.1 is estimated using method of moments. The system of equations presented in chapter 9 is solved using nonlinear least squares. MATLAB was used for the computation.[14] Functions from Michael Cliff's MINZ Program Libraries for MATLAB were used to solve the system of equations.[15]

Results are presented for the log-log functional form. Results are not presented for the log-linear functional form because they are nearly identical to the log-log functional form. School district coefficients are not presented. Any differences between the log-linear and log-log functional forms appear to be due to the using the predicted value versus the actual value of the land use indices. Recall in chapter 11 that the tests of endogeneity were similar for the log-linear and log-log functional forms, but not identical.

Table 12.1 provides a list of the variables, the abbreviations used in the remaining results tables, and a brief description.

[14] I would like to thank Mark Fleming for his invaluable assistance in using Cliff's functions.
[15] Functions were written using Cliff's routines. The author is happy to provide the functions and Monte Carlo tests results for their accuracy, if requested.

Variable Name Abbreviation	Variable Description
L_PRICE	Log of sale price
L_LOT	Log of lot size
L_SQFTSTR	Log of square foot of structure
BATH	Number of bathrooms
AC	Dummy for air conditioning
GARAGE	Dummy for garage
GOOD CONS	Good construction quality
VGOOD CONS	Very good construction quality
NUM STORY	Number of stories
FRAME	Dummy for frame construction
ALUM	Dummy of aluminum construction
1 Years Old	Age of structure is 1 year old
2-17 Years Old	Age of structure is 2 -17 years old
>54 Years Old	Age of structure is >54 years old
Summer Sale 1997	Sale took place between 04/01/97 and 10/1/97
Summer Sale 1998	Sale took place between 04/01/98 and 10/1/98
L_POPDEN	Log of population density
L_HSINCO	Log of median household income
L_DC	Log of distance to Washington, DC
L_TRAIN	Log of distance to mass transit train station
Viewshed PO	% permanent open space in viewshed
Neighborhood PO	% permanent open space in neighborhood
Viewshed DO	% developable open space in viewshed
Neighborhood DO	% developable open space in neighborhood
Viewshed LDENRES	% low density residential in viewshed
Neighborhood LDENRES	% low density residential in neighborhood
Viewshed HDENRES	% high density residential in viewshed
Neighborhood HDENRES	% high density residential in neighborhood

Table 12.1: Abbreviations and Descriptions of Variables Used in Final Model Estimation

12.2 Summarizing the Results of the Structural, Neighborhood, and Sales Period Variables

While the variables of interest are the land use indices, it is worthwhile to summarize the results from the other variables. This section begins by making comparisons of the parameter estimates across models estimated with different viewshed and neighborhood sizes and specifications for the spatial weight matrix. This section then discusses the parameter estimates themselves and interprets the results.

Table 12.2 presents the parameter estimates for the structural, sales period, and neighborhood variables. Table 12.2 is structured to allow comparisons of parameter estimates across the models hereafter defined by their viewshed-neighborhood combinations and two spatial weight matrices For instance, a comparison can be made between four parameter estimates for the variable L_LOT, given the Flexible 300-600 spatial weight matrix (or the Flexible Block-Sub-division) was used.

Table 12.2 reveals no instance of qualitative or quantitative difference when comparing the parameter estimates across viewshed-neighborhood combinations or spatial weight matrices. This is an expected result. The parameter estimates for the structural, sales period, and neighborhood variables should not be affected by changes in the viewshed-neighborhood combinations because they should be orthogonal to the land use indices, regardless of which indices are used. The parameter estimates for the structural, sales period, and neighborhood variables should not be affected by changes in the spatial weight matrix because the spatial weight matrices are similar in their structure.

	Flexible 300\600 Spatial Weight Matrix				Flexible Block\Sub-division Spatial Weight Matrix			
Buffer Size Combination	⅛-1	⅛- 1 ⅛	¼ - 1	¼ - 1 ½	⅛-1	⅛- 1 ½	¼ - 1	¼ - 1 ½
Variable								
CONSTANT	-0.0023	-0.0022	-0.0028	-0.0028	-0.0006	-0.0006	-0.0009	-0.0008
	(0.0037)	(0.0036)	(0.0036)	(0.0036)	(0.0029)	(0.0029)	(0.0029)	(0.0029)
L_LOT	**0.0977**	**0.0977**	**0.0945**	**0.0941**	**0.099**	**0.0989**	**0.0961**	**0.095**
	(0.006)	(0.006)	(0.0057)	(0.0059)	(0.006)	(0.0059)	(0.0057)	(0.0058)
L_SQFTSTR	**0.2958**	**0.2959**	**0.2951**	**0.2951**	**0.2997**	**0.2998**	**0.2982**	**0.2988**
	(0.0089)	(0.0089)	(0.0089)	(0.0089)	(0.0089)	(0.0089)	(0.0089)	(0.0089)
BATH	**0.0628**	**0.0626**	**0.0631**	**0.063**	**0.0623**	**0.0621**	**0.0624**	**0.0625**
	(0.0031)	(0.0031)	(0.0031)	(0.0031)	(0.0031)	(0.0031)	(0.0031)	(0.0031)
AC	**0.0516**	**0.0515**	**0.0517**	**0.0516**	**0.0479**	**0.048**	**0.0479**	**0.048**
	(0.0059)	(0.0058)	(0.0059)	(0.0059)	(0.0058)	(0.0058)	(0.0058)	(0.0058)
GARAGE	**0.055**	**0.0549**	**0.0553**	**0.0556**	**0.0555**	**0.0551**	**0.056**	**0.0559**
	(0.0052)	(0.0052)	(0.0052)	(0.0052)	(0.0052)	(0.0052)	(0.0052)	(0.0052)
GOOD CONS	**0.1294**	**0.1302**	**0.1277**	**0.1294**	**0.1316**	**0.1328**	**0.1293**	**0.1318**
	(0.0068)	(0.0067)	(0.0068)	(0.0068)	(0.0068)	(0.0068)	(0.0068)	(0.0068)
VGOOD	**0.2788**	**0.2798**	**0.2777**	**0.2792**	**0.279**	**0.2805**	**0.2781**	**0.2798**
CONS	(0.0111)	(0.0111)	(0.0111)	(0.0112)	(0.0111)	(0.0111)	(0.0111)	(0.0112)
NUM STORY	**0.0524**	**0.0526**	**0.0529**	**0.0532**	**0.0529**	**0.0532**	**0.0538**	**0.0541**
	(0.0061)	(0.0061)	(0.0061)	(0.0061)	(0.0061)	(0.0061)	(0.0061)	(0.0061)
FRAME	**-0.0205**	**-0.0208**	**-0.02**	**-0.0207**	**-0.0242**	**-0.0247**	**-0.023**	**-0.0243**
	(0.0055)	(0.0054)	(0.0055)	(0.0055)	(0.0055)	(0.0055)	(0.0055)	(0.0055)
ALUM	-0.0146	-0.0149	-0.0138	-0.0144	-0.0151	-0.0154	-0.0135	-0.0144
	(0.0079)	(0.0079)	(0.0079)	(0.0079)	(0.0079)	(0.0079)	(0.0079)	(0.0079)
1 YEAR OLD	**0.0734**	**0.0729**	**0.0711**	**0.0693**	**0.0748**	**0.074**	**0.0715**	**0.0692**
	(0.0103)	(0.0102)	(0.0103)	(0.0103)	(0.0103)	(0.0102)	(0.0103)	(0.0103)
2-17 YEARS	**0.0681**	**0.0678**	**0.0672**	**0.0667**	**0.073**	**0.0722**	**0.072**	**0.0708**
OLD	(0.0079)	(0.0079)	(0.0079)	(0.0079)	(0.0077)	(0.0077)	(0.0078)	(0.0078)
>54 YEARS	**0.022**	**0.0224**	**0.0226**	**0.0223**	**0.0267**	**0.0267**	**0.0269**	**0.0263**
OLD	(0.0077)	(0.0077)	(0.0077)	(0.0077)	(0.0078)	(0.0078)	(0.0079)	(0.0079)
SUMMER	0.0077	0.0076	0.008	0.008	0.0073	0.0071	0.0077	0.0077
SALE 1997	(0.0042)	(0.0042)	(0.0043)	(0.0042)	(0.0042)	(0.0042)	(0.0042)	(0.0042)
SUMMER	**0.0476**	**0.0476**	0.0476	**0.0475**	**0.0476**	**0.0477**	0.0477	**0.0475**
SALE 1998	(0.004)	(0.004)	(0.004)	(0.004)	(0.0039)	(0.0039)	(0.0039)	(0.0039)
L_POPDEN	-0.0062	-0.0079	-0.004	-0.0051	-0.0044	-0.0058	-0.001	-0.0023
	(0.0051)	(0.0051)	(0.0052)	(0.0052)	(0.0049)	(0.0048)	(0.0049)	(0.0049)
L_HSINCO	**0.1216**	**0.1216**	**0.1265**	**0.1244**	**0.1471**	**0.1476**	**0.1533**	**0.1519**
	(0.0115)	(0.0114)	(0.0113)	(0.0114)	(0.011)	(0.0109)	(0.0107)	(0.0109)
L_DC	**-0.1992**	**-0.1902**	**-0.1983**	**-0.1878**	**-0.2009**	**-0.1917**	**-0.2019**	**-0.1916**
	(0.0317)	(0.0326)	(0.0317)	(0.0326)	(0.0269)	(0.0277)	(0.0269)	(0.0277)
L_TRAIN	0.0062	0.0065	0.0102	0.0088	-0.0027	-0.0015	0.0012	0.0005
	(0.0083)	(0.0082)	(0.0078)	(0.0082)	(0.0071)	(0.007)	(0.0067)	(0.007)

Bold indicated significance at the 95% level for a two-tailed test

Table 12.2: Parameter Estimates on the Structural, Sales Period, and Neighborhood Variables—Eight Estimated Models Arranged by Buffer Size and Spatial Matrices

The coefficient estimates for the structural variables are consistent with intuition about structural attributes. **L_LOT**, **L_SQFTSTR**, **BATH**, and **NUM STORY** have

positive coefficient estimates, indicating that an additional unit of these attributes would increase the sale price of the house. The variables **AC** and **GARAGE** have positive coefficient estimates, indicating that the presence of these attributes increases the sale price. Variables **GOOD CONS** and **VGOOD CONS** also have positive coefficient estimates, indicating that higher construction quality relative to fair and poor increases the value of a home. Both **FRAME** and **ALUM** had negative coefficient values, suggesting that structures built with these materials have lower sale prices compared to structures built with other materials (such as brick and stucco). Coefficient estimates for the age groups reveal that new construction sells for a premium, compared to middle-age structures (17 - 54 years old) and older homes (more than 54 years) sell for a premium, compared to middle-age structures.

The results for the sale periods yield an unexpected result. While residential parcels sold during the summer sale period in 1998 had a price premium, the estimated coefficient for the summer sale period in 1997 is not significant at 95% level, suggesting no price premium. This may be caused by a down market in 1997, a boom in 1998, or by other factors relating to the supply of residential parcels for sale in 1998.[16]

The results for the neighborhood quality variables are mixed. The estimated coefficient for **L_POPDEN** (measured at the Census Block level in 2000) was not significant in any of the estimated models. The estimated parameters for **L_MEDHSINC** (measured at the Census Block level in 2000) are positive and significant, consistent with expectations. **L_MEDHSINC** could potentially be

[16] It was suggest to the author by a real estate agent at Long and Foster that the market was "hot" in the summer of 1998 (person communication).

endogenous with housing price. People with higher income tend to buy more expensive houses and conversely, people with expensive houses tend to have higher incomes.

The negative coefficient estimates for **L_DC** are consistent with expectations. The negative values indicate that as distance to Washington, D.C. increases, price declines. Without a doubt, people will pay a price premium to live closer to Washington, D.C. This is due to a host of factors, not the least of which is being closer to employment centers. On the other hand, the coefficient estimates for **L_TRAIN** are insignificant. Shorter distance to train as a form of public transportation was expected to increase the housing price. This result is puzzling and contrary to past work. One possible explanation is the type of neighborhoods near some Metro stations. The Glenmont, Wheaton, and Silver Spring neighborhoods, each of which contains Metro stops, are more "rough" neighborhoods than in the majority of Montgomery County.[17]

Although not presented in Table 12.2, the parameter estimates for 27 school districts were robust to changes in the viewshed-neighborhood combination and to changes in the specification of the spatial weight matrices.

12.3 The Robustness of the Land Use Index Parameter Estimates to Changes in the Spatial Weight Matrix Specification

It is useful to compare the coefficients for the land use indices between the two specifications for the spatial weight matrix. Parameter estimates for the land use indices between the two models are not expected to be qualitatively or quantitatively different

[17] I thank Economic Research Service Economists for pointing this out to me.

because the spatial weight matrices are similar in the number of neighbors. However, this is an empirical question not a model assumption.

	Viewshed-Neighborhood Combination							
	⅛-1		⅛- 1 ½		¼ - 1		¼ - 1 ½	
Spatial Weight Matrix	300/600	BLK/SUB	300/600	BLK/SUB	300/600	BLK/SUB	300/600	BLK/SUB
Variable								
Viewshed PO	**0.1051**	**0.1188**	**0.108**	**0.1223**	**0.0375**	**0.0443**	0.0279	**0.0398**
	(0.0265)	(0.0248)	(0.0263)	(0.0245)	(0.0154)	(0.0142)	(0.0197)	(0.017)
Neighborhood	**-0.0645**	**-0.0762**	**-0.2149**	**-0.2504**	**-0.1085**	**-0.1281**	**-0.2048**	**-0.2461**
PO	(0.0679)	(0.0539)	(0.0868)	(0.0704)	(0.0512)	(0.0405)	(0.0875)	(0.071)
Viewshed DO	**0.0581**	**0.0609**	**0.0596**	**0.063**	-0.1904	**-0.3205**	0.0268	0.0161
	(0.0293)	(0.0273)	(0.029)	(0.0271)	(0.1782)	(0.1529)	(0.0263)	(0.0229)
Neighborhood	-0.0212	0.003	-0.13	-0.1082	-0.065	-0.0429	-0.157	-0.1307
DO	(0.0796)	(0.0651)	(0.0926)	(0.0762)	(0.0635)	(0.052)	(0.0944)	(0.0779)
Viewshed	**0.0755**	**0.0821**	**0.0754**	**0.0821**	-0.082	-0.1536	-0.0305	-0.0405
LDENRES	(0.023)	(0.022)	(0.0229)	(0.0219)	(0.0939)	(0.0841)	(0.0251)	(0.0235)
Neighborhood	-0.0557	-0.0553	-0.1178	**-0.1447**	**-0.0922**	**-0.0986**	-0.0968	**-0.1272**
LDENRES	(0.0645)	(0.0519)	(0.079)	(0.0636)	(0.0409)	(0.0334)	(0.0792)	(0.0637)
Viewshed	0.0465	**0.0539**	0.0536	**0.0632**	-0.094	-0.1553	0.0011	0.0053
HDENRES	(0.0283)	(0.0264)	(0.0279)	(0.026)	(0.0912)	(0.0847)	(0.0239)	(0.0202)
Neighborhood	0.0858	0.0938	0.074	0.0477	0.2899	**0.3925**	0.064	0.0355
HDENRES	(0.0725)	(0.0586)	(0.0824)	(0.0666)	(0.1642)	(0.1355)	(0.0838)	(0.0678)

Bold indicates significance at the 95% level for a two-tailed test
Comparisons can be made between pairs of coefficients having the same viewshed- neighborhood combination, but different specifications for the spatial weight matrix

Table 12.3: Land Use Index Parameter Estimates--Ordered for Comparing Estimates between Models Using Two Different Spatial Weight Matrices Assumptions

The land use index parameter estimates are presented in Table 12.3, organized to make easy comparisons between two different specifications of the spatial weight matrix, but otherwise the same (same viewshed-neighborhood combination). Table 12.3 reveals very few instances of qualitative difference. Between the 24 pairs of coefficients, there are 7 instances of qualitative difference. There is no instance of quantitative difference between the 24 pairs of estimates. An additional conclusion from this table is that instances of qualitative difference were not restricted to any one model (any of the four viewshed-neighborhood combinations), i.e. no pattern of difference is found. Each of the

four viewshed-neighborhood combinations had at least one instance of qualitative difference in the parameter estimates.

Thus, the land use index parameter estimates are somewhat but not excessively sensitive to changes in the spatial weight matrix, with no one particular viewshed-neighborhood combination exhibiting more or less sensitivity to changes in the structure of the spatial weight matrix. However, both open space parameter estimates are much less sensitive than the residential density land use parameter estimates. Thus I conclude that the parameter estimates for open space are robust to changes in the specification of the spatial weight matrix

12.4 The Robustness of the Land Use Index Parameter Estimates to Changes in the Viewshed-Neighborhood Combinations

It is useful to compare the estimated coefficients for the land use indices between the four viewshed-neighborhood combinations. The expectation is that:

1) The parameter estimates for the viewshed land use indices should not be affected by changes to the neighborhood size
2) The parameter estimates for the neighborhood land use indices should not be affected by the changes to the viewshed size.

The two expectations are implicit in the model that is being estimated. The model is developed to test whether surrounding land uses impact the price of a residential parcel

at two levels: in the immediate area (viewshed) and the larger area (neighborhood). Although these actual percents of land uses measured in a viewshed and neighborhood may be correlated, the model assumes that they are independent of each other.

To address the two expectations, separate tables are constructed. These tables simply re-organize the results presented in Table 12.3 in a way that simplifies the confirmation of the hypothesis.

Table 12.4 addresses the expectation that the parameter estimates for the viewshed land use indices should not be affected by changes to the neighborhood size. Table 12.4 reveals that changing the neighborhood size has a small qualitative impact on the viewshed parameter estimates. There were two instances of qualitative difference between the parameter estimates (circled). Both of these were in cases where the larger viewshed size was used.

	Flexible 300\600 Spatial Weight Matrix				**Flexible Block\Sub-division Spatial Weight Matrix**			
Viewshed/ Neighborhood Combination	⅛-1	⅛- 1 ⅛	¼ - 1	¼ - 1 ½	⅛-1	⅛- 1 ½	¼ - 1	¼ - 1 ½
Variables								
Viewshed PO	**0.1051**	**0.108**	**0.0375**	0.0279	**0.1188**	**0.1223**	**0.0443**	**0.0398**
	(0.0265)	(0.0263)	(0.0154)	(0.0197)	(0.0248)	(0.0245)	(0.0142)	(0.017)
Viewshed DO	**0.0581**	**0.0596**	-0.1904	0.0268	**0.0609**	**0.063**	**-0.3205**	0.0161
	(0.0293)	(0.029)	(0.1782)	(0.0263)	(0.0273)	(0.0271)	(0.1529)	(0.0229)
Viewshed LDENRES	**0.0755**	**0.0754**	-0.082	-0.0305	**0.0821**	**0.0821**	-0.1536	-0.0405
	(0.023)	(0.0229)	(0.0939)	(0.0251)	(0.022)	(0.0219)	(0.0841)	(0.0235)
Viewshed HDENRES	0.0465	0.0536	-0.094	0.0011	**0.0539**	**0.0632**	-0.1553	0.0053
	(0.0283)	(0.0279)	(0.0912)	(0.0239)	(0.0264)	(0.026)	(0.0847)	(0.0202)

Bold indicates significance at the 95% level for a two-tailed test
Comparisons can be made between pairs of coefficients having the same size neighborhood

Table 12.4: Comparing the Viewshed Parameter Estimates between Models with Varying Levels of Neighborhood Size

Table 12.5 addresses the expectation that the parameter estimates for the neighborhood land use indices should not be affected by changes to the viewshed size. Table 12.5 reveals that changing the viewshed size has a small qualitative impact on the neighborhood parameter estimates. There were three instances of qualitative difference between the parameter estimates (circled). All three of the qualitative differences were in cases where the smaller neighborhood size was used.

	Flexible 300\600 Spatial Weight Matrix				Flexible Block\Sub-division Spatial Weight Matrix			
Viewshed/ Neighborhood Combination	⅛-1	¼ - 1	⅛- 1 ½	¼ - 1 ½	⅛-1	¼ - 1	⅛- 1 ½	¼ - 1 ½
Variables								
Neighborhood	**-0.0645**	**-0.1085**	**-0.2149**	**-0.2048**	**-0.0762**	**-0.1281**	**-0.2504**	**-0.2461**
PO	(0.0679)	(0.0512)	(0.0868)	(0.0875)	(0.0539)	(0.0405)	(0.0704)	(0.071)
Neighborhood	-0.0212	-0.065	-0.13	-0.157	0.003	-0.0429	-0.1082	-0.1307
DO	(0.0796)	(0.0635)	(0.0926)	(0.0944)	(0.0631)	(0.052)	(0.0762)	(0.0779)
Neighborhood	-0.0557	**-0.0922**	-0.1178	-0.0968	-0.0553	**-0.0986**	**-0.1447**	**-0.1272**
LDENRES	(0.0645)	(0.0409)	(0.079)	(0.0792)	(0.0519)	(0.0334)	(0.0636)	(0.0637)
Neighborhood	0.0858	0.2899	0.074	0.064	0.0938	**0.3925**	0.0477	0.0355
HDENRES	(0.0725)	(0.1642)	(0.0824)	(0.0838)	(0.0586)	(0.1355)	(0.0666)	(0.0678)

Comparisons can be made between pairs of coefficients having the same size viewshed

Table 12.5: Comparing the Neighborhood Parameter Estimates between Models with Varying Levels of Viewshed Size

12.5 Comparing the Results of the Spatial Dependence Coefficients

The hypothesis was made that spatial weight matrices built on bands based on geographic relationships within the blocks and sub-divisions may better capture spatial dependence in the errors of the regression. A flexible form for the weight matrix is proposed that allows potential spatial dependence with neighbors that are in a parcel's block level and within its subdivision but outside its block. This specification is

compared to a specification based on distance bands similar to Bell and Bockstael's. In this case, spatial dependence is allowed between parcels that within 300 meters from a parcels and then parcels between 300-600 meters.

Spatial dependence at the block level is expected to be higher than that at the sub-division level. In fact, results from the spatial weight matrix based on bands of distance set below 300 meters and between 300 and 600 meters corroborate Bell and Bockstael's findings that spatial dependence decreases as distance increases. In every model that was estimated (including log-linear), spatial dependence based on the 300 meter band is larger than that on the 300 - 600 meter band.

The results based on the flexible form of the spatial weight matrix using blocks and sub-divisions reveal the opposite result. Spatial dependence based on the block level is smaller than that at the sub-division level. This is completely opposite of what was expected. However, it should be noted that hypothesis tests for the spatial dependence parameters can not be carried out, so it is not clear if the estimates are significantly different from zero. A different formulation of blocks could be more appropriate.

Model fit was also compared between models based on both specifications of the spatial weight matrix. Model fit was evaluated based on value of the log likelihood, evaluated at the GM estimates for lambda and the FGLS estimates for σ^2. In all four models (four different viewshed-neighborhood combinations) the 300/600 flexible had a better fit than the Block/Sub-division flexible form.

To illustrate why this is occurring, Table 9.4 is presented again. This table shows that, on average, the number of neighbors at the block level is one third less than the number of neighbors within 300 meters. However, the number of neighbors between 300

and 600 meters is less than the number of neighbors in the area outside the block but within the sub-division.

Name	Mean	Std. Dev.	Min	Max
CONTIG 300	16.12	14.12	1	115
CONTIG 300-600	27.91	18.27	1	115
CONTIG BLK	5.52	6.42	1	40
CONTIG BLK-SUB	30.53	28.41	1	146

Table 9.4: Statistics on Number of Neighbors in Each Specification of the Spatial Weight Matrix

The variables included in the model could explain price well at the very local level with limited remaining spatial dependence at that level. Or it could be that unobservable variables are not as correlated at the block level as they are at the sub-division level.

12.6 Conclusions

Chapter 12 sought to determine how robust the parameter estimates were to changes in the specification of the spatial weight matrix and to changes in the viewshed-neighborhood combination. The following conclusions were reached:

1) The structural, neighborhood, time, and school district dummies were robust to changes in the specification of the spatial weight matrix and viewshed-neighborhood combination.

2) Land use index parameter estimates were fairly robust to changes in the specification of the spatial weight matrix.

3) Parameter estimates for land use indices measured at the viewshed level were robust to changes in the size of the neighborhood

4) Parameter estimates for land use indices measured at the neighborhood were robust to changes in the size of the viewshed

5) Spatial dependence appears to increase from the block level to the sub-division.

6) The 300/600 specification for the spatial weight matrix had a better model fit than the Block/sub-division specification across all viewshed-neighborhood combinations.

Chapter 13

INTERPRETING THE LANDUSE PARAMETER ESTIMATES AND SIMULATING LOCALIZED INCREASES IN PERMANENTLY PRESERVED OPEN SPACE

13.1 Overview

The purpose of this thesis is to examine 1) if adjacent landowners receive benefits in terms of a higher parcel value from adjacent open space. Another question was whether these benefits result in additional property taxes that are sufficiently large to finance the preservation of additional agricultural land. Chapters 11 and 12 provide some evidence on how the answers to these two questions may be sensitive to the methodology use to answer them. This chapter answers both of these questions in the context of the results presented in the previous two chapters. The second section of this chapter focuses on comparing the values of permanently preserved and developable open space, paying special attention to how the results hold across different models. The third section of this chapter describes the simulation process used to determine if the benefits to nearby property owners of permanently preserved open space are enough to finance preservation programs. The fourth section of this chapter discusses the results of the simulation.

13.2 Comparing the Land Use Index Parameter Estimates

The coefficient estimates for the land use indices variables can be interpreted as elasticities. The model uses a log-log functional form. The land use indices are measured as percentages and are not logged. However, they can still be interpreted as quasi-elasticities in the sense that a change of 1 percent of a land use index will cause a β percent change in the dependent variable.

The conclusion reached in section 12.3 is that the open space parameter estimates were robust to changes in the specification of the spatial weight matrix. In addition, in all models, the 300/600 specification of the spatial weight matrix have a better fit that the Block/Sub-division specification. Therefore, the results from the 300/600 models are used to make the comparison of the values of different types of open space across different viewshed-neighborhood combinations.

Table 13.1 presents the results from the four viewshed-neighborhood combinations using the 300/600 specification of the spatial weight matrix. All else being equal, an increase in permanently preserved open space around a parcel has a positive and significant impact on the value of that residential parcel at the viewshed level. This result does not hold for the ¼ mile-1 ½ mile viewshed-neighborhood combination, where the parameter estimate for land use at the ¼ mile is not significantly different from zero.

	Viewshed/ Neighborhood combination			
	⅛-1	⅛- 1 ½	¼ - 1	¼ - 1 ½
Variable				
Viewshed PO	**0.1051**	**0.108**	**0.0375**	0.0279
	(0.0265)	(0.0263)	(0.0154)	(0.0197)
Neighborhood	**-0.0645**	**-0.2149**	**-0.1085**	**-0.2048**
PO	(0.0679)	(0.0868)	(0.0512)	(0.0875)
Viewshed DO	**0.0581**	**0.0596**	-0.1904	0.0268
	(0.0293)	(0.029)	(0.1782)	(0.0263)
Neighborhood	-0.0212	-0.13	-0.065	-0.157
DO	(0.0796)	(0.0926)	(0.0635)	(0.0944)
Viewshed	**0.0755**	**0.0754**	-0.082	-0.0305
LDENRES	(0.023)	(0.0229)	(0.0939)	(0.0251)
Neighborhood	-0.0557	-0.1178	**-0.0922**	-0.0968
LDENRES	(0.0645)	(0.079)	(0.0409)	(0.0792)
Viewshed	0.0465	0.0536	-0.094	0.0011
HDENRES	(0.0283)	(0.0279)	(0.0912)	(0.0239)
Neighborhood	0.0858	0.074	0.2899	0.064
HDENRES	(0.0725)	(0.0824)	(0.1642)	(0.0838)

Bold indicates significance at the 95% level for a two-tailed test

Table 13.1: Land Use Index Parameter Estimates for 300/600 Specification of the Spatial Weight Matrix

Comparing the permanently preserved open space parameter estimates to the developable open space estimates shows that in three out of four instances, an increase in permanently preserved open space around a parcel increases housing price more than an increase in developable open space. This is also true for low-density residential and high-density residential land uses. This suggests that the conversion of developable open space to permanently preserved open space increases the value of a residential parcel. This result holds in both cases of the ⅛ mile viewshed and one case of the ¼ mile viewshed, where the estimate for permanently preserved open space was positive and the estimate for developable open space was not significantly different from zero at the 95% level. In the other case of the ¼ mile viewshed, the parameter estimates for both

permanently preserved open space and developable open space are not significantly different than zero

The permanently preserved open space parameter estimates for the neighborhood are found to be negative and significant, regardless of the viewshed-neighborhood combination. This is a curious result and will be discussed further in the third section of this chapter.

All four developable open space parameter estimates in the neighborhood are found to be negative, but not significantly different than zero. The parameter estimates for low-density residential are not significantly different than zero in three out of four models, and negative and significant in the fourth case. The parameter estimates for high-density residential are positive, but not significantly different than zero in all four viewshed-neighborhood combinations. This suggests that the value of residential parcels is not affected by the density of residential housing nearby, all else the same.

13.3 Simulation Process

The previous section addressed the question of adjacent landowners receiving benefits in terms of a higher parcel value from adjacent open space. The results were mixed, finding increased value in the viewshed and decreased value in the neighborhood. The remaining question is whether these benefits (if any) result in additional property taxes which are sufficiently large to finance the preservation of additional agricultural land. In order to calculate the net benefits of an increase in preserved open space to

surrounding residential parcel owners, a simulation is conducted. The purpose of this section is to detail the simulation process.

The first step in this process is to select parcels (or pieces of parcels) that are currently (1998) in developable open space and can be converted to permanently preserved open space. For this simulation, two parcels are selected. The parcels are selected with the intention of simulating a conversion in both an urban part of the County and a sub-urban/rural part of the County. To simulate an urban area conversion, a parcel in Olney was chosen. To simulate a conversion in a suburban/rural area, a parcel in Poolesville is chosen. The characteristics of Olney and Poolesville will be discussed in the next section.

The second step in the process is to count the total number of residential parcels, both sample and non-sample, whose viewsheds and neighborhoods contain any part of the newly preserved parcel. This yields the number of residential parcels that would be impacted from an increase in permanently preserved open space, or, all the residential parcels whose viewsheds and neighborhoods would realize an increase of permanently preserved open space.

The third step is to calculate the average sales price of the residential parcels within each of the four distances from the newly preserved parcel. Since all sample and non-sample residential parcels must be counted, and there is no sales data on the non-sample parcels, an assumption is made that the average price of all the residential parcels is equal to the average price of the residential parcels in the sample.

The fourth step is to determine the amount of increase (in percent terms) of permanently preserved open space from the newly preserved open space parcel. The

newly preserved parcel will result in an increase of permanently preserved open space within the viewshed and neighborhood but the increase will be different for individual residential parcels, given it depends on how much of the newly preserved parcel is contained within each parcel's viewshed or neighborhood. Therefore, the average increase for all residential parcels (sample and non-sample) is calculated, within each viewshed and neighborhood.

The fifth step is to derive a coefficient for the expected change in housing value for each viewshed-neighborhood combination. It is assumed that an increase of permanently preserved open space will come from a decrease in developable open space. Therefore, to get the net value (net elasticity), the developable open space coefficient must be subtracted from the permanently preserved open space coefficient.

The sixth step is to calculate the total value of the increase in permanently preserved open space. This is achieved by multiplying the number of homes, average price, average percent increase in open space, and the net value coefficient.

The seventh step is to calculate the present value of the future stream of tax revenues from an increase in taxable value, if one is found. A discount rate is set at 5% with a period of 30 years. The discount rate and time horizon are chosen ad hoc.

The eight and final step is to compare the present value of the future stream of tax revenues to the cost of converting the parcel to permanently preserved open space (i.e., easement cost). The cost of an easement is determined by an Easement Value Worksheet and will range from $800 to $4500 per acre.

13.4 Results of Simulations

As mentioned in the previous sections, two parcels of developable open space were chosen for the simulation that derives the net value of converting a parcel to permanently preserved open space.

The first simulation is a conversion to permanently preserved open space of an 11 acre parcel of developable open space in the city of Olney. Olney is a city of approximately 32,000 persons (2000 Census). It has a population density of 2,420 persons per square mile, compared with 1,762 for all of Montgomery County. On average, it has a younger population than Montgomery County. The median household income in Olney in 2000 was $94,818. Olney is located approximately 13 miles from the center of Washington, DC and is generally regarded as a bedroom community for Washington.

Table 13.2 shows that the netl value of a conversion of 11 acres of developable open space to permanently preserved open space to nearby residential parcel owners is negative. The value to residential parcel owners in the viewshed is positive, but this positive value is overshadowed by the negative value measured in the neighborhood. The difference between the positive and negative values varies, depending on the viewshed-neighborhood combination, but the take home message is the same: The conversion of a parcel of developable open space to permanently preserved open space has a net negative effect on the price of residential parcels within a 1 ½ mile radius of it. To some extent, this is due to the fact that the larger buffers contain more residential properties

Viewshed/ Neighborhood	Step 2 Number of Residential Parcels	Step 3 Average Price of Residential Parcels	Step 4 Avg % Open Space Added	Step 5 Net Coeff.	Step 6 Total Change in Residential Parcel Value (in millions)
1/8 - 1 mile					
Viewshed	99	$233,120	6.980	0.0440	$7.1
Neighborhood	1504	$242,734	0.485	-0.0645	-$11.4
Net value					-$4.3
1/8 - 1.5 mile					
Viewshed	99	$233,120	6.980	0.0484	$7.8
Neighborhood	2757	280,587	0.234	-0.2149	-$38.0
Net value					-$31.2
1/4 - 1 mile					
Viewshed	260	$233,855	4.357	0.0375	$9.9
Neighborhood	1504	$242,734	0.485	-0.1085	-$19.2
Net value					-$9.3
1/4 - 1.5 mile					
Viewshed	260	$242,734	4.357	0.0000*	$0
Neighborhood	2757	$280,587	0.485	-0.2048	-$76.8
Net value					-$76.8

*Coefficient not significantly different from zero for both developable and permanently preserved open space

Table 13.2 Net Value of Conversion of 11 Acre Parcel to Permanently Preserved Open Space in Olney, Maryland

The second simulation is a conversion to permanently preserved open space of an 8 acre parcel of developable open space in the town of Poolesville. Poolesville is a city of approximately 5,151 persons (2000 Census). It has a population density of 1,333 persons per square mile compared with 1,762 for all of Montgomery County. On average, it has a younger population than Montgomery County. The median household

income in Poolesville in 2000 was $85,091. Poolesville is regarded as a bedroom community for Washington, D.C.

Viewshed/ Neighborhood	Step 2 Number of Residential Parcels	Step 3 Average Price of Residential Parcels	Step 4 Avg % Open Space Added	Step 5 Net Coeff.	Step 6 Total Change in Residential Parcel Value (in millions)
1/8 - 1 mile					
Viewshed	53	$310,790	6.530	0.0440	$4.7
Neighborhood	483	$229,202	0.485	-0.0645	-$3.4
Net value					$1.3
1/8 - 1.5 mile					
Viewshed	53	$310,790	6.530	0.0484	$5.2
Neighborhood	677	$236,111	0.234	-0.2149	-$8.0
Net value					-$2.8
1/4 - 1 mile					
Viewshed	118	$306,495	3.780	0.0375	$5.1
Neighborhood	483	229,202	0.485	-0.1085	-$5.8
Net value					-$.7
1/4 - 1.5 mile					
Viewshed	118	$306,495	3.780	0.0000*	$0
Neighborhood	677	$236,111	0.485	-0.2048	-$15.9
Net value					-$15.9

*Coefficient not significantly different from zero for both developable and permanently preserved open space

Table 13.3 Net Value of Conversion of an 8 Acre Parcel to Permanently Preserved Open Space in Poolesville, Maryland

Table 13.3 shows that the net value of a conversion of 8 acres of developable open space to permanently preserved open space to all residential parcel owners within 1 ½ miles is negative, but may be positive within 1 mile. The value to residential parcel owners in the viewshed is positive, but negative when measured in the neighborhood.

The difference between the positive and negative values is not consistent across viewshed-neighborhood combinations. In the case of ⅛ mile - 1 mile viewshed-neighborhood combination, the positive value measured at the viewshed is greater than the negative value measured at the neighborhood. However, when the neighborhood size is increased to 1 ½ miles, leaving the viewshed the same, the positive value measured at the viewshed is exceeded by the negative value measured at the neighborhood. In the two instances where the viewshed is measured at ¼ mile, the negative value when measured at the neighborhood is larger than the value measured at the viewshed, resulting in a net overall loss of residential value.

Given that a positive net value is found at the ⅛ - 1 mile viewshed-neighborhood combination, the next step is to calculate the present value of future tax revenue from the increase in taxable value of the residential parcels. The total net value (viewshed - neighborhood) is $1.27 million per year. Montgomery County's property tax rate is $.754 for every $100 of taxable value. This results in $9575 additional tax revenue every year. At a discount rate of 5%, discounted over 30 years, the present value of the conversion is $154551. The cost of an easement in Montgomery County ranges from $800 to $4500 per acre. Therefore, the cost of an 8 acre easement would range from $6400 to $36000. The present value of the future stream of tax revenue is more than four times the highest possible cost of the easement. Therefore, the conversion of developable open space to permanently preserved open space pays for itself over time, in this case.

The most interesting result from the two simulations is that the value of permanently preserved open space is negative when measured at the neighborhood level, regardless of the size of the neighborhood. This is a curious result given the expectation

that the conversion of developable open space to permanently preserved open space should increase the value of residential parcels. One important note is that both developable and permanently preserved open space coefficients, measured at the neighborhood level, were negative. The fact that a open space parcel was preserved or developable did not have an opposite effect at the neighborhood level.

In similar previous research by Geoghegan, Lynch, and Bucholtz (2003), the effect of open space on nearby residential parcels was found to be negative in Carroll County. In Howard and Calvert County, the effects were positive but not found for both the viewshed and/or neighborhood. Therefore, in the context of previous research, the findings from the simulation for Montgomery County are not surprising.

The simulations demonstrate that the value of open space to nearby residential parcel owners is generally positive when measured in the viewshed and generally negative when measured in the neighborhood. However, the net effect was negative in most cases. The measurement is sensitive to the viewshed-neighborhood combination and to the location where the simulation was conducted.

Chapter 14

CONCLUSION

The purpose of this thesis was to determine if proximity to permanently preserved open space increased the value of nearby residential parcels in Montgomery County, Maryland. The value of permanently preserved open space was compared to developable open space to determine their relative worth. The comparison was made for two distinct areas around a residential parcel meant to reflect a parcel's viewshed and general neighborhood. Lastly, a simulation was conducted to calculate the net value to nearby residential parcel owners of a conversion of developable open space to permanently preserved open space. This total value was used to determine if the present value of the future stream of increased tax revenues accruing due to any positive increases in housing price from the conversion to permanently preserved open space could finance the preservation activity. To conduct the analysis, a hedonic model was developed that included land use indices. In conducting this analysis, three methodological issues were addressed. Endogeneity was found to exist in the land use indices when measured in the larger viewshed. Its existence raised questions for future research.

The second methodological issue that was addressed was spatial dependence in the error term of the regression model. It was found that spatial dependence existed. Spatial dependence was modeled in the error term. It was posited that the spatial relationship among residential parcel is more accurately portrayed using the city block and sub-divisions to define "neighbors" rather than distance alone. This was compared to

a more general form using distance bands. It was found that the parameter estimates were not sensitive to the changes in how spatial dependence was modeled (i.e., changes in the structure of the spatial weight matrix). However, the level of spatial dependence was found to be sensitive to the specification of the structure of spatial dependence. Spatial dependence was found to be greater among residential parcels within the sub-division but outside one's block when compared to parcels within one's block. This result was opposite of expectations and past findings. The take home message is that the total value of open space is robust to changes in the specification of spatial dependence among the residential parcels.

The last methodological issue addressed was changes in the size of the viewshed and neighborhood for which the land use indices were measured. The results suggested that an increase in permanently preserved open space increases the value of residential parcels in the viewshed, but decreases that value of residential parcels in the neighborhood. The net value, that is, the sum of the viewshed and neighborhood values, was generally negative, but sensitive to the size of the land use indices. In some instances, the net value may be positive, while other instances it may be negative. Where the net value of an increase in open space was positive, it was determined that the cost of preserving the developable open space land through the county-level PDR program was less than the present value of the increase in the future stream of tax revenues from nearby residential parcel. The take home message is that several sizes for the viewshed and neighborhood should be used to determine the robustness of the total value of an increase in permanently preserved open

REFERENCES

Anselin, L. 1988. *Spatial Econometrics: Methods and Models.* Dordrecht: Kluwer Academic Publishers.

Anselin, L., and Griffith, Daniel A. 1988. "Do Spatial Effects Really Matter in Regression Analysis?" *Papers of the Regional Science Association* 65: 11-34

Anselin, L. 2002. "Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models." *Agricultural Economics* 27(3): 247-67.

Bell, K.P. and N.E. Bockstael. 2000. "Applying the Generalized Method of Moments Estimation Approach to Spatial Problems Involving Microlevel Data." *Review of Economics and Statistic* 82(1): 72-82.

Bockstael, N. E. 1996. "Modeling Economics and Ecology: The Importance of a Spatial Perspective." *American Journal of Agricultural Economics* 78 (5): 1168-80.

Bolduc, D., Laferrierre, R., and Santarossa, G. 1992. "Spatial Autoregressive Error Components in Travel Flow Models." *Regional Science and Urban Economics* 22: 381-385.

Can, Ayse. 1992. "Specification and Estimation of Hedonic Housing Price Models." *Regional Science and Urban Economics* 22: 453-474.

Case, A.C. 1991. "Spatial Patterns in Household Demand." *Econometrica*, 59: 953-965.

Case, A.C. 1992. "Neighborhood Influence and Technological Change." *Regional Science and Urban Economics*, 22: 491-508.

Chattopadhyay, S. 1999. "Estimating the Demand for Air Quality. New Evidence Based on the Chicago Housing Market." *Land Economics* 75(1): 1-22.

Chesapeake Bay Foundation. 2002. *Future Growth in the Washington, D.C. Metropolitan Area, Annapolis, Maryland.*

Chesapeake Bay Commission and The Trust for Public Land. 2001. *Keeping Our Commitment: Preserving Land in the Chesapeake Watershed.* Annapolis, MD.

Cliff, A. and Ord, J.K. 1973. *Spatial Autocorrelation*. Pion, London.

Correll, M.R., J.H. Lillydahl, and L.D. Singell. 1978. "The Effects of Greenbelts on Residential Property Values: Some Findings on the Political Economy of Open Space." *Land Economics* 54 (May): 207-17.

Espey, M. and H. Lopez. 2000. "The Impact of Airport Noise on Willingness to Pay for Residences. *Growth and Change* 31: 408-419.

Garrod, G. and K. Willis. 1992a. "The Amenity Value of Woodland in Great Britain: A Comparison of Economic Estimates." *Environmental and Resource Economics* 2: 415-34.

_____. 1992b. "The Environmental Economic Impact of Woodland: A Two Stage Hedonic Price Model of the Amenity Value of Forestry in Britain." *Applied Economics* 24 (7): 715-28.

_____. 1992c. "Valuing Goods' Characteristics: An Application of the Hedonic Price Method to Environmental Attributes." *Journal of Environmental Management* 34: 59-76.

Geoghegan, J. 2002 "The Value of Open Spaces in Residential Land Use." *Land Use Policy*, Vol. 19, No. 1, pp. 91-98.

Geoghegan, J., Lynch, L., and S. J. Bucholtz. 2003. "Capitalization of Open Space into Housing Values: Can Agricultural Easement Programs Pay for Themselves?" Agricultural and Resource Economic Review 32(1): 33-45

Geoghegan, J., L. Wainger, & N. E. Bockstael. 1997. "Spatial Landscape Indices in a Hedonic Framework: An Ecological Economics Analysis Using GIS." *Ecological Economics* 23 (3): 251-64.

Green, W. H. *Econometric Analysis*. 2000. Fourth Ed, Prentice Hall, NJ.

Hammer, T.R., R.E. Coughlin, and E.T. Horn IV. 1974. "The Effect of a Large Urban Park on Real Estate Values." *American Institute of Planners Journal* (July): 274-77.

Irwin, E. G. and N. E. Bockstael. 2001. "The Problem of Identifying Land Use Spillovers: Measuring the Effects of Open Space on Residential Property Values," *American Journal of Agricultural Economics*, 83(3): 698-704.

Irwin, E. G. (2002). "The Effects of Open Space on Residential Property Values." *Land Economics* 78 (Nov): 465-80.

Kitchen, J.W., and W.S. Hendon. 1967. "Land Values Adjacent to An Urban Neighborhood Park." *Land Economics* 43 (Aug): 357-61.

Kelejian, H. H. and Prucha, I. 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model." *International Economic Review* 40: 509-533.

Kohlhase, J. E. 1991. "The Impact of Toxic Waste Sites on Housing Values." *Journal of Urban Economics* 30: 1-26.

Leggett, C.G. and N.E. Bockstael, 2000. "Evidence of the Effects of Water Quality on Residential Land Prices." *Journal of Environmental Economics and Management* 39(2): 121-144.

LeSage J. P., and R. Kelley Pace. 2000. "Using Matrix Exponentials to Explore Spatial Structure in Regression Relationships." Working Paper.

Lima, E. C. R., and P. B. R. Macedo. 1999. "Estimation of a Weights Matrix for Determining Spatial Effects." Working Paper 672.

Lynch, L., and S. J. Lovell, 2002. "Hedonic Price Analysis of Easement Payments in Maryland Agricultural Land Preservation Programs, " Department of Agricultural and Resource Economics Working Paper #02-13, 2002 University of Maryland, College Park.

McMillan, M. 1974. "Open Space Preservation in Developing Areas: An Alternative Policy." *Land Economics* 50 (Nov): 410-17.

Michaels, R. G., and V. K. Smith. 1990. "Market Segmentation and Valuing Amenities with Hedonic Models: The Case of Hazardous Waste Sites." *Journal of Urban Economics* 28: 223-242.

Nickerson, C., and L. Lynch. 2001. "The Effect of Farmland Preservation Programs on Farmland Prices." *American Journal of Agricultural Economics* 83(2): 341-351.

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, 2003.

Pace, R. K., and R. Barry. 1997. ''Sparse Spatial Autoregressions,'' *Statistics and Probability Letters* 33: 291–297.

Pace, R. K., R. Barry, V.C. Slawson, and C.F. Sirmans. 2000. "Simultaneous Spatial and Function Form Transformations," *Advances in Spatial Econometrics*, edited by Raymond Florax and Luc Anselin.

Palmquist, R. B. 1992. "Valuing Localized Externalities." *Journal of Urban Economics* 31: 59-68.

Peiser, R.B., and G.M. Schwann. 1993. "The Private Value of Public Open Space Within Sub-divisions." *Journal of Architectural and Planning Research* 10 (Summer): 91-104.

Ridker, R. G. and J. A. Henning. 1967. "The Determinants of Residential Property Values with Special Reference to Air Pollution." *The Review of Economics and Statistics* 49: 246-257.

Rosen, S. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82(1): 35-55.

Tyrävinen, L. 1997. "The Amenity Value of the Urban Forest: An Application of the Hedonic Pricing Method."" *Landscape and Urban Planning* 37: 211-222.

Tyrävinen, L, and A. Meittinen. 2000. "Property Prices and Urban Forest Amenities." *Journal of Environmental Economics and Management* 39(2): 205-223.

Weicher, J.C., and R.H. Zerbst. 1973. "The Externalities of Neighborhood Parks: An Empirical Investigation." *Land Economics* 49 (Feb): 99-105.